

# On self-similar sets with overlaps and inverse theorems for entropy

Michael Hochman

## Abstract

We study the Hausdorff dimension of self-similar sets and measures on  $\mathbb{R}$ . We show that if the dimension is smaller than the minimum of 1 and the similarity dimension, then at small scales there are super-exponentially close cylinders. This is a step towards the folklore conjecture that such a drop in dimension is explained only by exact overlaps, and confirms the conjecture in cases where the contraction parameters are algebraic. It also gives an affirmative answer to a conjecture of Furstenberg, showing that the projections of the “1-dimensional Sierpinski gasket” in irrational directions are all of dimension 1.

As another consequence, when a family of self-similar sets or measures is parametrized in a real-analytic manner, then, under an extremely mild non-degeneracy condition, the set of “exceptional” parameters has Hausdorff dimension 0. Thus, for example, there is at most a zero-dimensional set of parameters  $1/2 < \lambda < 1$  such that the corresponding Bernoulli convolution has dimension  $< 1$ , and similarly for Sinai’s problem on iterated function systems that contract on average.

A central ingredient of the proof is an inverse theorem for the growth of entropy of convolutions of probability measures. For the dyadic partition  $\mathcal{D}_n$  of  $\mathbb{R}$  into cells of side  $2^{-n}$ , we show that if  $\frac{1}{n}H(\nu * \mu, \mathcal{D}_n) \leq \frac{1}{n}H(\mu, \mathcal{D}_n) + \delta$ , then, when restricted to random element of a partition  $\mathcal{D}_i$ ,  $1 \leq i \leq n$ , either  $\mu$  is close to uniform or  $\nu$  is close to atomic. This should be compared to results in additive combinatorics that give the global structure of measures satisfying  $\frac{1}{n}H(\nu * \mu, \mathcal{D}_n) \leq \frac{1}{n}H(\mu, \mathcal{D}_n) + O(\frac{1}{n})$ .

## 1 Introduction

### 1.1 Self-similar sets and measures

In this paper an iterated function system (IFS) will mean a finite family  $\Phi = \{\varphi_i\}_{i \in \Lambda}$  of linear contractions of  $\mathbb{R}$ ,  $\varphi_i(x) = r_i x + a_i$  with  $|r_i| < 1$  and  $a_i \in \mathbb{R}$ . To avoid trivialities we assume throughout that there are at least two distinct contractions. A self similar set is the attractor of such a system, i.e. the unique compact set  $\emptyset \neq X \subseteq \mathbb{R}$  satisfying

$$X = \bigcup_{i \in \Lambda} \varphi_i X. \quad (1)$$

The self-similar measure associated to a probability vector  $(p_i)_{i \in \Lambda}$  is the unique Borel probability measure  $\mu$  on  $\mathbb{R}^d$  satisfying

$$\mu = \sum_{i=1}^k p_i \cdot \varphi_i \mu. \quad (2)$$

---

Supported by ERC grant 306494

2010 *Mathematics Subject Classification*. 28A80, 11K55, 11B30, 11P70

Here  $\varphi\mu = \mu \circ \varphi^{-1}$  denotes the push-forward of  $\mu$  by  $\varphi$ .

When the images  $\varphi_i X$  are disjoint or satisfy various weaker separation assumptions, the small-scale structure of these objects is well understood, and in particular the Hausdorff dimension  $\dim X$  of  $X$  is equal to the similarity dimension<sup>1</sup>  $\text{s-dim } X$ , i.e. the unique solution  $s \geq 0$  of the equation  $\sum |r_i|^s = 1$ . Defining the dimension of a measure  $\theta$  by<sup>2</sup>

$$\dim \theta = \inf\{\dim E : \theta(E) > 0\}$$

and assuming again sufficient separation of the images  $\varphi_i X$ , the dimension  $\dim \mu$  of a self-similar measure  $\mu$  is equal to the similarity dimension of  $\mu$ , defined by

$$\text{s-dim } \mu = \frac{\sum p_i \log p_i}{\sum p_i \log r_i}.$$

When the images  $\varphi_i X$  have significant overlap, however, far less is known about the structure, or even the dimension, of these objects. One can give trivial bounds: the dimension is never greater than the similarity dimension, and it is never greater than the dimension of the ambient space  $\mathbb{R}$ , which is 1. Hence

$$\dim X \leq \min\{1, \text{s-dim } X\} \quad (3)$$

$$\dim \mu \leq \min\{1, \text{s-dim } \mu\}. \quad (4)$$

However, without special combinatorial assumptions on the IFS, current methods are unable even to decide whether or not equality holds in (3) and (4), let alone compute the dimension exactly. The exception is when there are sufficiently many exact overlaps among the “cylinders” of the IFS. More precisely, for  $i = i_1 \dots i_n \in \Lambda^n$  write

$$\varphi_i = \varphi_{i_1} \circ \dots \circ \varphi_{i_n}.$$

One says that exact overlaps occur if there is an  $n$  and distinct  $i, j \in \Lambda^n$  such that  $\varphi_i = \varphi_j$  (in particular the images  $\varphi_i X$  and  $\varphi_j X$  coincide).<sup>3</sup> If this occurs then  $X$  and  $\mu$  can be expressed using an IFS  $\Psi$  which is a proper subset of  $\{\varphi_i\}_{i \in \Lambda^n}$ , and a strict inequality in (3) and (4) may follow from the corresponding bound for  $\Psi$ .

## 1.2 Main results

This work was motivated by the folklore conjecture that *the occurrence of exact overlaps is the only mechanism which can lead to a strict inequality in (3) and (4)* (see e.g. [24, question 2.6]). Our main result lends some support to the conjecture and proves some special cases of it. All of our results hold, with suitable modifications, in higher dimensions, but this will appear separately.

Fix  $\Phi = \{\varphi_i\}_{i \in \Lambda}$  as in the previous section and for  $i \in \Lambda^n$  write  $r_i = r_{i_1} \dots r_{i_n}$ , which is the contraction ratio of  $\varphi_i$ . Define the distance between the cylinders associated

<sup>1</sup>This notation is imprecise, since the similarity dimension depends on the IFS  $\Phi$  rather than the attractor  $X$ , but the meaning should always be clear from the context. A similar remark holds for the similarity dimension of measures.

<sup>2</sup>This is the lower Hausdorff dimension. There are many other notions of dimension but for self-similar measures all the major ones coincide since such measures are exact dimensional [8].

<sup>3</sup>One should note that if  $i \in \Lambda^k$ ,  $j \in \Lambda^m$  and  $\varphi_i = \varphi_j$ , then  $i$  cannot be a prefix of  $j$  and vice versa, so  $ij, ji \in \Lambda^{k+m}$  are distinct and  $\varphi_{ij} = \varphi_{ji}$ . This shows that our definition of exact overlaps includes coincidence of cylinders at “different generations”.

to  $i, j \in \Lambda^n$  by

$$d(i, j) = \begin{cases} \infty & r_i \neq r_j \\ |\varphi_i(0) - \varphi_j(0)| & r_i = r_j \end{cases}.$$

Note that  $d(i, j) = 0$  if and only if  $\varphi_i = \varphi_j$  and that the definition is unchanged if 0 is replaced by any other point. For  $n \in \mathbb{N}$  let

$$\Delta_n = \min\{d(i, j) : i, j \in \Lambda^n, i \neq j\}.$$

Exact overlaps occur if and only if  $\Delta_n = 0$  for some  $n$  (equivalently all sufficiently large  $n$ ). One also always has exponential decay of  $\Delta_n$ . Indeed, all of the points  $\varphi_i(0)$ ,  $i \in \Lambda^n$ , belong to a fixed bounded interval (independent of  $n$ ), and the exponentially many sequences  $i \in \Lambda^n$  give rise to only polynomially many contraction ratios  $r_i$ . Therefore there are distinct  $i, j \in \Lambda^n$  with  $r_i = r_j$  and  $|\varphi_i(0) - \varphi_j(0)| < |\Lambda|^{-(1-o(1))n}$ , which implies that  $\Delta_n \rightarrow 0$  exponentially. In general this is all one can say, since in many cases there is an exponential lower bound for  $\Delta_n$ . Such a lower bound occurs when the images  $\varphi_i(X)$  are disjoint but also sometimes when they intersect, for instance in Garsia's example from [11] and the examples considered in Theorems 1.5 and 1.6 below.

**Theorem 1.1.** *If  $\mu$  is a self-similar measure on  $\mathbb{R}$  and if  $\dim \mu < \min\{1, \text{s-dim } \mu\}$ , then  $\Delta_n \rightarrow 0$  super-exponentially, i.e.  $\lim(-\frac{1}{n} \log \Delta_n) = \infty$ .*

Note that the conclusion is in terms of the sequence  $\Delta_n$ , which is determined by the IFS  $\Phi$ , not the measure. Thus if the conclusion fails, the hypothesis must fail for all self-similar measures of  $\Phi$ . Every self-similar set  $X$  supports a self-similar measure  $\mu$  with  $\text{s-dim } \mu = \text{s-dim } X$ , and since we always have  $\dim \mu \leq \dim X$ , we conclude:

**Corollary 1.2.** *If  $X$  is the attractor of an IFS on  $\mathbb{R}$  and if  $\dim X < \min\{1, \text{s-dim } X\}$ , then  $\lim(-\frac{1}{n} \log \Delta_n) = \infty$ .*

Theorem 1.1 is derived from a more quantitative result about the entropy of finite approximations of  $\mu$ . Write  $H(\mu, \mathcal{E})$  for the Shannon entropy of a measure  $\mu$  with respect to a partition  $\mathcal{E}$ , and  $H(\mu, \mathcal{E}|\mathcal{F})$  for the conditional entropy on  $\mathcal{F}$ ; see Section 3.1. For  $n \in \mathbb{Z}$  the dyadic partitions of  $\mathbb{R}$  into intervals of length  $2^{-n}$  is

$$\mathcal{D}_n = \{[\frac{k}{2^n}, \frac{k+1}{2^n}) : k \in \mathbb{Z}\}.$$

For  $t \in \mathbb{R}$  we also write  $\mathcal{D}_t = \mathcal{D}_{[t]}$ . We remark that  $\liminf \frac{1}{n} \log H(\theta, \mathcal{D}_n) \geq \dim \theta$  for any probability measure  $\theta$ , and the limit exists and is equal to  $\dim \theta$  when  $\theta$  is exact dimensional, which is the case for self-similar measures [8].

We first consider the case that  $\Phi$  is uniformly contracting, i.e. that all  $r_i$  are equal to some fixed  $r$ . Fix a self-similar measure  $\mu$  defined by some probability vector  $(p_i)_{i \in \Lambda}$  and for  $i \in \Lambda^n$  write  $p_i = p_{i_1} \cdots p_{i_n}$ . Without loss of generality one can assume that 0 belongs to the attractor  $X$ . Define the  $n$ -th generation approximation of  $\mu$  by

$$\nu^{(n)} = \sum_{i \in \Lambda^n} p_i \cdot \delta_{\varphi_i(0)}. \quad (5)$$

This is a probability measure on  $X$  and  $\nu^{(n)} \rightarrow \mu$  weakly. Moreover, writing

$$n' = n \log_2(1/r),$$

$\nu^{(n)}$  closely resembles  $\mu$  up to scale  $2^{-n'} = r^n$  in the sense that

$$\lim_{n \rightarrow \infty} \frac{1}{n'} H(\nu^{(n)}, \mathcal{D}_{n'}) = \dim \mu.$$

The main question we are interested in is the behavior of  $\nu^{(n)}$  at smaller scales. Observe that the entropy  $H(\nu^{(n)}, \mathcal{D}_{n'})$  of  $\nu^{(n)}$  at scale  $2^{-n'}$  may not exhaust the entropy  $H(\nu^{(n)})$  of  $\nu^{(n)}$  as a discrete measure (i.e. with respect to the partition into points). If there is substantial excess entropy it is natural to ask at what scale and at what rate it appears; it must appear eventually because  $\lim_{k \rightarrow \infty} H(\nu^{(n)}, \mathcal{D}_k) = H(\nu^{(n)})$ . The excess entropy at scale  $k$  relative to the entropy at scale  $n'$  is just the conditional entropy  $H(\nu^{(n)}, \mathcal{D}_k | \mathcal{D}_{n'}) = H(\nu^{(n)}, \mathcal{D}_k) - H(\nu^{(n)}, \mathcal{D}_{n'})$ .

**Theorem 1.3.** *Let  $\mu$  be a self-similar measure on  $\mathbb{R}$  defined by an IFS with uniform contraction ratios. Let  $\nu^{(n)}$  be as above. If  $\dim \mu < 1$ , then*

$$\lim_{n \rightarrow \infty} \frac{1}{n'} H(\nu^{(n)}, \mathcal{D}_{qn'} | \mathcal{D}_{n'}) = 0 \quad \text{for every } q > 1. \quad (6)$$

We now formulate the result in the non-uniformly contracting case. Let

$$r = \prod_{i \in \Lambda} r_i^{p_i}$$

so that  $\log r$  is the average logarithmic contraction ratio when  $\varphi_i$  is chosen randomly with probability  $p_i$ . Note that, by the law of large numbers, with probability tending to 1, an element  $i \in \Lambda^n$  chosen according to the probabilities  $p_i$  will satisfy  $r_i = r^{n(1+o(1))} = 2^{n'(1+o(1))}$ .

With this definition and  $\nu^{(n)}$  defined as before, the theorem above holds as stated, but note that now the partitions  $\mathcal{D}_k$  are not suitable for detecting exact overlaps, since  $\varphi_i(0) = \varphi_j(0)$  may happen for some  $i, j \in \Lambda^n$  with  $r_i \neq r_j$ . To correct this define the probability measure  $\tilde{\nu}^{(n)}$  on  $\mathbb{R} \times \mathbb{R}$  by

$$\tilde{\nu}^{(n)} = \sum_{i \in \Lambda^n} \delta_{(\varphi_i(0), r_i)}$$

and the partition of  $\mathbb{R} \times \mathbb{R}$  given by

$$\tilde{\mathcal{D}}_n = \mathcal{D}_n \times \mathcal{F},$$

where  $\mathcal{F}$  is the partition of  $\mathbb{R}$  into points.

**Theorem 1.4.** *Let  $\mu$  be a self-similar measure on  $\mathbb{R}$  and  $\tilde{\nu}^{(n)}$  as above. If  $\dim \mu < 1$ , then*

$$\lim_{n \rightarrow \infty} \frac{1}{n'} H(\tilde{\nu}^{(n)}, \tilde{\mathcal{D}}_{qn'} | \tilde{\mathcal{D}}_{n'}) = 0 \quad \text{for every } q > 1. \quad (7)$$

To derive Theorem 1.1, let  $\mu$  be as in the last theorem with  $\dim \mu < \min\{1, \text{s-dim } \mu\}$ . The conclusion of the last theorem is equivalent to  $\frac{1}{n'} H(\mu, \tilde{\mathcal{D}}_{qn'}) \rightarrow \dim \mu$  for every  $q > 1$ . Hence for a given  $q$  and all sufficiently large  $n$  we will have  $\frac{1}{n'} H(\tilde{\nu}^{(n)}, \tilde{\mathcal{D}}_{qn'}) < \text{s-dim } \mu$ . Since  $\tilde{\nu}^{(n)} = \sum_{i \in \Lambda^n} p_i \cdot \delta_{(\varphi_i(0), r_i)}$ , if each pair  $\varphi_i(0), \varphi_j(0)$  in the sum belonged to a different atom of  $\tilde{\mathcal{D}}_{qn'}$  then we would have  $\frac{1}{n'} H(\tilde{\nu}^{(n)}, \tilde{\mathcal{D}}_{qn'}) = -\frac{1}{n \log(1/r)} \sum_{i \in \Lambda^n} p_i \log p_i = \text{s-dim } \mu$ , a contradiction. Thus there must be distinct  $i, j \in \Lambda^n$  for which  $\varphi_i(0), \varphi_j(0)$  lie in the same atom of  $\tilde{\mathcal{D}}_{qn'}$ , giving  $\Delta_n < 2^{-qn'}$ .

### 1.3 Outline of the proof

Let us say a few words about the proofs. For simplicity we discuss Theorem 1.3, where there is a common contraction ratio  $r$  to all the maps. For a self similar measure  $\mu = \sum_{i \in \Lambda} p_i \cdot \varphi_i \mu$ , iterate the relation  $n$  times to get  $\mu = \sum_{i \in \Lambda^n} p_i \cdot \varphi_i \mu$ . Since each  $\varphi_i$ ,  $i \in \Lambda^n$ , contracts by  $r^n$ , all the measures  $\varphi_i \mu$ ,  $i \in \Lambda^n$ , are translates of each other, so this can be re-written as a convolution

$$\mu = \nu^{(n)} * \tau^{(n)},$$

where as before  $\nu^{(n)} = \sum_{i \in \Lambda^n} p_i \cdot \delta_{\varphi_i(0)}$ , and  $\tau^{(n)}$  is a translate of  $\mu$  scaled down by  $r^n$ .

Fix  $q$  and write  $a \approx b$  to indicate that the difference tends to 0 as  $n \rightarrow \infty$ . From the entropy identity  $H(\mu, \mathcal{D}_{(q+1)n'}) = H(\mu, \mathcal{D}_{n'}) + H(\mu, \mathcal{D}_{(q+1)n'} | \mathcal{D}_{n'})$  and the fact that  $H(\mu, \mathcal{D}_{n'}) \approx H(\nu^{(n)}, \mathcal{D}_{n'})$ , we find that the mean entropy

$$A = \frac{1}{(q+1)n'} H(\mu, \mathcal{D}_{(q+1)n'})$$

is approximately a convex combination  $A \approx \frac{1}{(q+1)}B + \frac{q}{(q+1)}C$  of the mean entropy

$$B = \frac{1}{n'} H(\nu^{(n)}, \mathcal{D}_{n'})$$

and the mean conditional entropy

$$C = \frac{1}{qn'} H(\mu, \mathcal{D}_{(q+1)n'} | \mathcal{D}_{n'}) = \sum_{I \in \mathcal{D}_{n'}} \mu(I) \cdot \frac{1}{qn'} H(\nu_I^{(n)} * \tau^{(n)}, \mathcal{D}_{(q+1)n'}),$$

where  $\nu_I^{(n)}$  is the conditional measure of  $\nu^{(n)}$  on  $I$ . Since  $A \approx \dim \mu$  and  $B \approx \dim \mu$ , we find that  $C \approx \dim \mu$  as well. On the other hand we also have  $\frac{1}{qn'} h(\tau^{(n)}, \mathcal{D}_{(q+1)n'}) \approx \dim \mu$ , thus

$$\frac{1}{qn'} H(\nu_I^{(n)} * \tau^{(n)}, \mathcal{D}_{(q+1)n'}) = C \approx \dim \mu \approx \frac{1}{qn'} H(\tau^{(n)}, \mathcal{D}_{(q+1)n'}) \quad (8)$$

for large  $n$  and typical  $I \in \mathcal{D}_{n'}$ . The argument is then concluded by showing that (8) implies that either  $\frac{1}{qn'} H(\tau^{(n)}, \mathcal{D}_{(q+1)n'}) \approx 1$  (leading to  $\dim \mu = 1$ ), or that typically  $\frac{1}{qn'} H(\nu_I^{(n)}, \mathcal{D}_{(q+1)n'}) \approx 0$  (leading to (6)).

Now, for a general pair of measures  $\nu, \tau$  the relation  $\frac{1}{k} H(\nu * \tau, \mathcal{D}_k) \approx \frac{1}{k} H(\nu, \mathcal{D}_k)$  analogous to (8) does not have such an implication. But, while we know nothing about the structure of  $\nu_I^{(n)}$ , we do know that  $\tau^{(n)}$ , being self-similar, is highly uniform at different scales. We will be able to utilize this fact to draw the desired conclusion. Evidently, the main ingredient in the argument is an analysis of the growth of measures under convolution, which will occupy us starting in Section 2.

### 1.4 Applications

Theorem 1.1 and its corollaries settle a number of cases of the aforementioned conjecture. Specifically, in any class of IFSs where one can prove that cylinders are either equal or exponentially separated, the only possible cause of dimension drop is the occurrence of exact overlaps. Thus,

**Theorem 1.5.** *For IFSs on  $\mathbb{R}$  defined by algebraic parameters, there is a dichotomy: Either there are exact overlaps or the attractor  $X$  satisfies  $\dim X = \min\{1, \text{s-dim } X\}$ .*

*Proof.* Let  $\varphi_i(x) = r_i x + a_i$  and suppose  $r_i, a_i$  are algebraic. For distinct  $i, j \in \Lambda^n$  the distance  $|\varphi_i(0) - \varphi_j(0)|$  is a polynomial of degree  $n$  in  $r_i, a_i$ , and hence is either equal to 0, or is  $\geq s^n$  for some constant  $s > 0$  depending only on the numbers  $r_i, a_i$  (see Lemma 5.10). Thus  $\Delta_n \geq s^n$  and the conclusion follows from Corollary 1.2.  $\square$

There are a handful of cases where a similar argument can handle non-algebraic parameters. Among these is a well-known conjecture by Furstenberg from the 1970s on the linear images of the “one dimensional Sierpinski gasket”. Let  $F \subseteq \mathbb{R}^2$  be given by

$$F = \left\{ \sum (i_n, j_n) 3^{-n} : (i_n, j_n) \in \{(0, 0), (1, 0), (0, 1)\} \right\},$$

and let  $\pi_t : \mathbb{R}^2 \rightarrow \mathbb{R}$  denote the linear map

$$\pi_t(x, y) = tx + y.$$

Then  $F_t = \pi_t F$  is a self-similar subset of  $\mathbb{R}$  defined by the contractions

$$x \mapsto \frac{1}{3}x \quad , \quad x \mapsto \frac{1}{3}(x+1) \quad , \quad x \mapsto \frac{1}{3}(x+t). \quad (9)$$

Furstenberg conjectured that  $\dim \pi_t F = 1$  for all irrational  $t$  (see e.g. [24, question 2.5]). To relate this to our main conjecture, note that  $\text{s-dim } F_t = 1$  for all  $t$  and that exact overlaps occur only for certain rational values of  $t$ . From general considerations such as Marstrand’s theorem, we know that  $\dim F_t = 1$  for a.e.  $t$ , and Kenyon showed that this holds also for a dense  $G_\delta$  set of  $t$  [18]. In the same paper Kenyon also classified those rational  $t$  for which  $\dim F_t = 1$ , and showed that  $F_t$  has Lebesgue measure 0 for all irrational  $t$  (strengthening the conclusion of a general theorem of Besicovitch that gives this for a.e.  $t$ ). For some other partial results see also [30].

**Theorem 1.6.** *If  $t \notin \mathbb{Q}$  then  $\dim F_t = 1$ .*

*Proof.* Fix  $t$ , and suppose that  $\dim F_t < 1$ . Let  $\Lambda = \{0, 1, t\}$  and  $\varphi_i(x) = x/3 + i$ , so  $F_t$  is the attractor of  $\{\varphi_i\}_{i \in \Lambda}$ . Let  $X_n = \{\sum_{i=1}^n a_i 3^{-i} : a_i \in \{\pm 1, 0\}\}$ . For each  $n$  and  $i, j \in \Lambda^n$  we have  $|\varphi_i(0) - \varphi_j(0)| = p_{i,j} - t \cdot q_{i,j}$  for some  $p_{i,j}, q_{i,j} \in X_n$ , so there are  $p_n, q_n \in X_n$  such that  $\Delta_n = |p_n - t \cdot q_n|$ . Now, by Corollary 1.2,  $|p_n - t \cdot q_n| = \Delta_n < 30^{-n}$  for all large enough  $n$ , which, since  $q_n > 3^{-n}$ , gives  $|t - p_n/q_n| < 10^{-n}$ . Subtracting successive terms, by the triangle inequality we have

$$\left| \frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n} \right| < 2 \cdot 10^{-n} \quad \text{for large enough } n.$$

But  $p_n, q_n, p_{n+1}, q_{n+1} \in X_n$ , so  $p_{n+1}/q_{n+1} - p_n/q_n$  is rational with denominator  $\geq 4 \cdot 9^{-n}$ , giving

$$\left| \frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n} \right| \neq 0 \implies \left| \frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n} \right| \geq 4 \cdot 9^{-n}.$$

For large  $n$  the last two inequalities are incompatible unless  $p_n/q_n = p_{n+1}/q_{n+1}$ . In other words, there is an  $n_0$  such that  $|t - p_{n_0}/q_{n_0}| < 10^{-n}$  for  $n > n_0$  which gives  $t = p_0/q_0$ .  $\square$

The argument above is due to B. Solomyak and P. Shmerkin and we thank them for permission to include it here. Similar considerations work in a few other cases, but one already runs into difficulties if in the example above we replace the contraction ratio  $1/3$  with a general non-algebraic  $0 < r < 1$  (see also the discussion following Theorem 1.9 below).

In the absence of a resolution of the general conjecture, we turn to parametric families of self-similar sets and measures. The study of parametric families of general sets and measures is classical; examples include the projection theorems of Besicovitch and Marstrand and more recent results like those of Peres-Schlag [21] and Bourgain [3]. When the sets and measures in question are self-similar we shall see that the general results can be strengthened considerably.

Let  $I$  be a set of parameters, let  $r_i : I \rightarrow (-1, 1) \setminus \{0\}$  and  $a_i : I \rightarrow \mathbb{R}$ ,  $i \in \Lambda$ . For each  $t \in I$  define  $\varphi_{i,t} : \mathbb{R} \rightarrow \mathbb{R}$  by  $\varphi_{i,t}(x) = r(t)(x - a_i(t))$ . For a sequence  $i \in \Lambda^n$  let  $\varphi_{i,t} = \varphi_{i_1,t} \circ \dots \circ \varphi_{i_n,t}$  and define

$$\Delta_{i,j}(t) = \varphi_{i,t}(0) - \varphi_{j,t}(0). \quad (10)$$

The quantity  $\Delta_n = \Delta_n(t)$  associated as in the previous section to the IFS  $\{\varphi_{i,t}\}_{i \in \Lambda}$  is not smaller than the minimum of  $|\Delta_{i,j}(t)|$  over distinct  $i, j \in \Lambda^n$  (since it is the minimum over pairs  $i, j$  with  $r_i = r_j$ ). Thus,  $\Delta_n \rightarrow 0$  super-exponentially implies that  $\min\{|\Delta_{i,j}(t)|, i, j \in \Lambda^n, i \neq j\} \rightarrow 0$  super-exponentially as well, so Theorem 1.1 has the following formal implication.

**Theorem 1.7.** *Let  $\Phi_t = \{\varphi_{i,t}\}$  be a parametrized IFS as above. For every  $\varepsilon > 0$  let*

$$E_\varepsilon = \bigcup_{N=1}^{\infty} \bigcap_{n>N} \left( \bigcup_{i,j \in \Lambda^n} (\Delta_{i,j})^{-1}(-\varepsilon^n, \varepsilon^n) \right) \quad (11)$$

and

$$E = \bigcap_{\varepsilon>0} E_\varepsilon. \quad (12)$$

Then for  $t \in I \setminus E$ , for every probability vector  $p = (p_i)$  the associated self-similar measure  $\mu_t$  of  $\Phi_t$  satisfies  $\dim \mu_t = \min\{1, \text{s-dim } \mu_t\}$  and the attractor  $X_t$  of  $\Phi_t$  satisfies  $\dim X_t = \min\{1, \text{s-dim } X_t\}$ .

Our goal is to show that the set  $E$  is a small. We restrict ourselves to the case that  $I \subseteq \mathbb{R}$  is a compact interval; a multi-parameter version will appear in [13]. Extend the definition of  $\Delta_{i,j}$  to infinite sequences  $i, j \in \Lambda^{\mathbb{N}}$  by

$$\Delta_{i,j}(t) = \lim_{n \rightarrow \infty} \Delta_{i_1 \dots i_n, j_1 \dots j_n}(t). \quad (13)$$

Convergence is uniform over  $I$  and  $i, j$ , and if  $a_i(\cdot)$  and  $r(\cdot)$  are real analytic, so are the functions  $\Delta_{i,j}(\cdot)$ .

**Theorem 1.8.** *Let  $I \subseteq \mathbb{R}$  be a compact interval, let  $r : I \rightarrow (-1, 1) \setminus \{0\}$  and  $a_i : I \rightarrow \mathbb{R}$  be real analytic, and let  $\Phi_t = \{\varphi_{i,t}\}_{i \in \Lambda}$  be the associated parametric family of IFSs, as above. Suppose that*

$$\forall i, j \in \Lambda^{\mathbb{N}} \quad (\Delta_{i,j} \equiv 0 \text{ on } I \iff i = j).$$

Then the set  $E$  of “exceptional” parameters in Theorem 1.7 has Hausdorff and packing dimension 0.

Most existing results on parametric families of IFSs are based on the so-called transversality method, introduced by Pollicott and Simon [25] and developed, among others, by Solomyak [29] and Peres-Schlag [21]. Theorem 1.8 is based on a similar but much weaker “higher order” transversality condition, which is automatically satisfied under the stated hypothesis. We give the details in Section 5.4. See [28] for an effective derivation of higher-order transversality in certain contexts.

As a demonstration we apply this to the Bernoulli convolutions problem. For  $0 < \lambda < 1$  let  $\nu_\lambda$  denote the distribution of the real random variable  $\sum_{n=0}^{\infty} \pm \lambda^n$ , where the signs are chosen i.i.d. with equal probabilities. The name derives from the fact that  $\nu_\lambda$  is the infinite convolution of the measures  $\frac{1}{2}(\delta_{-\lambda^n} + \delta_{\lambda^n})$ ,  $n = 0, 1, 2, \dots$ , but the pertinent fact for us is that  $\nu_\lambda$  is a self-similar measure, given by assigning equal probabilities to the contractions

$$\varphi_{\pm}(x) = \lambda x \pm 1. \quad (14)$$

For  $\lambda < \frac{1}{2}$  the measure is supported on a self-similar Cantor set of dimension  $< 1$ , but for  $\lambda \in [\frac{1}{2}, 1)$  the support is an interval, and it is a longstanding problem to determine whether it is absolute continuous. Exact overlaps can occur only for certain algebraic  $\lambda$ , and Erdős showed that when  $\lambda^{-1}$  is a Pisot number  $\nu_\lambda$  is in fact singular [5]. No other parameters  $\lambda \in [\frac{1}{2}, 1)$  are known for which  $\nu_\lambda$  is singular. In the positive direction, it is known that  $\nu_\lambda$  is absolutely continuous for a.e.  $\lambda \in [1/2, 1)$  (Solomyak [29]) and the set of exceptional  $\lambda \in [a, 1)$  has dimension  $< 1 - C(a - 1/2)$  for some  $C > 0$  (Peres-Schlag [21]) and its dimension tends to 0 as  $a \rightarrow 1$  (Erdős [6]).

We shall consider the question of when  $\dim \nu_\lambda = 1$ . This is weaker than absolute continuity but little more seems to be known about this question except the relatively soft fact that the set of parameters with  $\dim \nu_\lambda = 1$  is also topologically large (contains a dense  $G_\delta$  set); see [22]. In particular the only parameters  $\lambda \in [1/2, 1)$  for which  $\dim \nu_\lambda < 1$  is known are inverses of Pisot numbers (Alexander-Yorke [1]). We also note that in many of the problems related to Bernoulli convolutions it is the dimension of  $\nu_\lambda$ , rather than its absolute continuity, that are relevant. For discussion of some applications see [22, Section 8] and [26].

**Theorem 1.9.**  *$\dim \nu_\lambda = 1$  outside a set of  $\lambda$  of dimension 0.*

*Proof.* Take the parametrization  $r(t) = t$ ,  $a_{\pm}(t) = \pm 1$  for  $t \in [1/2, 1 - \varepsilon]$ . Then  $\Delta_{i,j}(t) = \sum (i_n - j_n) \cdot t^n$  and this vanishes identically if and only if  $i = j$ , confirming the hypothesis of Theorem 1.8.  $\square$

Arguing as in the proof of Theorem 1.6, in order to show that  $\dim \nu_\lambda = 1$  for all non-algebraic  $\lambda$ , it would suffice to answer the following question in the affirmative:

**Question 1.10.** *Let  $\Pi_n$  denote the collection of polynomial of degree  $\leq n$  with coefficients  $0, \pm 1$ . Does there exist a constant  $s > 0$  such that for  $\alpha, \beta$  that are roots of polynomials in  $\Pi_n$  either  $\alpha = \beta$  or  $|\alpha - \beta| > s^n$ ?*

Classical bounds imply that this is true if  $s$  grows linearly in  $n$ , but we have not found an answer to the question in the literature.

Another problem to which our methods apply is the Keane-Smorodinsky  $\{0, 1, 3\}$ -problem. For details about the problem we refer to Pollicott-Simon [25] or Keane-Smorodinsky-Solomyak [17].



Finally, our methods also can be adapted with minor changes to IFSs that “contract on average” [20]. We restrict attention to a problem raised by Sinai [23] concerning the maps  $\varphi_- : x \mapsto (1 - \alpha)x - 1$  and  $\varphi_+ : x \mapsto (1 + \alpha)x + 1$ . A composition of  $n$  of these maps chosen i.i.d. with probability  $\frac{1}{2}, \frac{1}{2}$  asymptotically contracts by approximately  $(1 - \alpha^2)^{n/2}$ , and so for each  $0 < \alpha < 1$  there is a unique probability measure  $\mu_\alpha$  on  $\mathbb{R}$  satisfying  $\mu_\alpha = \frac{1}{2}\varphi_- \mu_\alpha + \frac{1}{2}\varphi_+ \mu_\alpha$ . Little is known about the dimension or absolute continuity of  $\mu_\alpha$  beyond upper bounds analogous to (4). Some results in a randomized analog of this model have been obtained by Peres, Simon and Solomyak [23]. We prove

**Theorem 1.11.**  $\dim \mu_\alpha = \min\{1, \text{s-dim } \mu_\alpha\}$  for  $\alpha \in (0, 1)$  outside a set of Hausdorff (and packing) dimension 0.

For further discussion of this problem see Section 5.5.

## 1.5 Absolute continuity?

There is another well-known conjecture, analogous to the one we started with, about the absolute continuity of self-similar measures  $\mu$  satisfies  $\text{s-dim } \mu > 1$ . Specifically, it has been suggested that such measures should be absolutely continuous as long as there are no exact overlaps. The Bernoulli convolutions problem discussed above is a special case of this conjecture.

Our methods at present are not able to address this. At a technical level, whenever our methods give  $\dim \mu = 1$  it is a consequence of showing that  $H(\mu, \mathcal{D}_n) = n - o(n)$ . In contrast, absolute continuity would require better asymptotics, e.g.  $H(\mu, \mathcal{D}_n) = n - O(1)$  (see [12, Theorem 1.5]). More substantially, our arguments do not distinguish between the critical ( $\text{s-dim } \mu = 1$ ) and super-critical ( $\text{s-dim } \mu > 1$ ) phases, so in their present form they cannot possibly give results about absolute continuity.

## 1.6 Notation and organization of the paper

The main ingredient in the proofs are our results on the growth of convolutions of measures. We develop this subject in the next three sections: Section 2 introduces the statements and basic definitions, Section 3 contains some preliminaries on entropy and convolutions, and Section 4 proves the the main results on convolutions. In Section 5 we prove Theorem 1.1 and the other main results.

We follow standard notational conventions.  $\mathbb{N} = \{1, 2, 3, \dots\}$ . All logarithms are to base 2.  $\mathcal{P}(X)$  is the space of probability measures on  $X$ , endowed with the weak-\* topology if appropriate. We follow standard “big  $O$ ” notation:  $O_\alpha(f(n))$  is an unspecified function bounded in absolute value by  $C \cdot f(n)$  for some constant  $C = C(\alpha)$  depending on  $\alpha$ . Similarly  $o(1)$  is a quantity tending to 0 as the relevant parameter  $\rightarrow \infty$ . The statement “for all  $s$  and  $t > t(s), \dots$ ” should be understood as saying “there exists a function  $t(\cdot)$  such that for all  $s$  and  $t > t(s), \dots$ ”. If we want to refer to a specific bound after the context where it is introduced we will designate it as  $t_1(\cdot)$ ,  $t_2(\cdot)$ ,  $t_*(\cdot)$ , etc.

## Acknowledgment

I am grateful to Pablo Shmerkin and Boris Solomyak for many contributions which have made this a better paper, and especially their permission to include the derivation of Theorem 1.6. I also thank Nicolas de Saxce for his comments. This project began

during a visit to Microsoft Research in Redmond, Washington, and I would like to thank Yuval Peres and the members of the theory group for their hospitality.

## 2 An inverse theorem for the entropy of convolutions

### 2.1 Entropy and additive combinatorics

As we saw at the end of Section 1.2, a key ingredient in the proof of Theorems 1.3 is an analysis of the growth of measures under convolution. This subject is of independent interest and will occupy us for a large part of this paper.

It will be convenient to introduce the normalized scale- $n$  entropy

$$H_n(\mu) = \frac{1}{n} H(\mu, \mathcal{D}_n).$$

Our aim is to obtain structural information about measures  $\mu, \nu$  for which  $\mu * \nu$  is small in the sense that

$$H_n(\mu * \nu) \leq H_n(\mu) + \delta, \tag{15}$$

where  $\delta > 0$  is small but fixed, and  $n$  is large.

This problem is a relative of classical ones in additive combinatorics concerning the structure of sets  $A, B$  whose sumset  $A + B = \{a + b : a \in A, b \in B\}$  is appropriately small. The general principle is that when the sum is small, the sets should have some algebraic structure. Such results are known as inverse theorems. For example the Freiman-Ruzsa theorem asserts that if  $|A + B| \leq C|A|$  then  $A, B$  are close (in a manner depending on  $C$ ) to (generalized) arithmetic progressions (the converse is immediate).<sup>4</sup> For details and more discussion see e.g [32].

With regard to entropy, in a recent paper Tao [31] obtained analogs of Freiman's theorem for the entropy of discrete measures, showing essentially that if

$$H_n(\mu * \mu) \leq H_n(\mu) + O\left(\frac{1}{n}\right) \tag{16}$$

then  $\mu, \nu$  are close, in an appropriate sense, to uniform measures on (generalized) arithmetic progressions.

The condition (15), however, is much weaker than (16) and it is harder to draw conclusions from it about the global structure of  $\mu$  (although some information can be obtained using the asymmetric Balog-Szemerédi-Gowers theorem). Consider the following example. Start with an arithmetic progression of length  $n_1$  and gap  $\varepsilon_1$ , and put the uniform measure on it. Now split each atom  $x$  into an arithmetic progression of length  $n_2$  and gap  $\varepsilon_2 < \varepsilon_1/n_2$ , starting at  $x$  (so the entire gap fits in the space between  $x$  and the next atom). Repeat this procedure  $N$  times with parameters  $n_i, \varepsilon_i$ , and call the resulting measure  $\mu$ . Let  $k$  be such that  $\varepsilon_N$  is of order  $2^{-k}$ . It is not hard to verify that we can have  $H_k(\mu) = 1/2$  but  $|H_k(\mu) - H_k(\mu * \mu)|$  arbitrarily small. This example is actually a (generalized) arithmetic progression, as predicted by Freiman-type theorems, but the rank  $N$  can be arbitrarily large. Furthermore if one conditions  $\mu$  on an exponentially small subset of its support one gets another example with the similar properties that is quite far from a generalized arithmetic progression.

Our main contribution to this matter is Theorem 2.7 below, which shows that constructions like the one above are, in a sense, the only way that (15) can occur. We

---

<sup>4</sup>A generalized arithmetic progression is an affine image of a box in a higher-dimensional lattice.

note that there is a substantial existing literature on the growth condition  $|A + B| \leq |A|^{1+\delta}$ , which is the sumset analog of 15. Such a condition appears in the sum-product theorems of Bourgain-Katz-Tao [4] and the work of Katz-Tao [16], and in the Euclidean setting more explicitly in Bourgain's work on the Erdős-Volkmann conjecture [2] and Marstrand-like projection theorems [3]. However we have not found a result in the literature that meets our needs and, in any event, we believe that the formulation given here will find further applications.

## 2.2 Component measures

The following notation will be needed in  $\mathbb{R}^d$  as well as  $\mathbb{R}$ . Let  $\mathcal{D}_n^d = \mathcal{D}_n \times \dots \times \mathcal{D}_n$  denote the dyadic partition of  $\mathbb{R}^d$ ; we often suppress the superscript when it is clear from the context. Let  $\mathcal{D}_n(x) \in \mathcal{D}_n$  denote the unique level- $n$  dyadic cell containing  $x$ . For  $D \in \mathcal{D}_n$  with  $\mu(D) > 0$ , let  $T_D : \mathbb{R}^d \rightarrow \mathbb{R}^d$  be the unique homothety mapping  $D$  to  $[0, 1)^d$ , and  $T_D\mu$  the push-forward of  $\mu$  through  $T_D$ .

**Definition 2.1.** For  $\mu \in \mathcal{P}(\mathbb{R}^d)$  and a dyadic cell  $D$  with  $\mu(D) > 0$ , the (raw)  $D$ -component of  $\mu$  is

$$\mu_D = \frac{1}{\mu(D)}\mu|_D$$

and the (rescaled)  $D$ -component is

$$\mu^D = \frac{1}{\mu(D)}T_D(\mu|_D).$$

For  $x \in \mathbb{R}^d$  with  $\mu(\mathcal{D}_n(x)) > 0$  we write

$$\begin{aligned}\mu_{x,n} &= \mu_{\mathcal{D}_n(x)} \\ \mu^{x,n} &= \mu^{\mathcal{D}_n(x)}.\end{aligned}$$

Taken together as  $x$  ranges over the support of  $\mu$ , these are the level- $n$  components of  $\mu$ .

Our results on the multi-scale structure of  $\mu \in \mathbb{R}^d$  are stated in terms of the behavior of random components of  $\mu$ , defined as follows.<sup>5</sup>

**Definition 2.2.** Let  $\mu \in \mathcal{P}(\mathbb{R}^d)$ .

1. A random level- $n$  component, raw or rescaled, is the random measure  $\mu_D$  or  $\mu^D$ , respectively, obtained by choosing  $D \in \mathcal{D}_n$  with probability  $\mu(D)$ ; equivalently, the random measure  $\mu_{x,n}$  or  $\mu^{x,n}$ , respectively, with  $x$  chosen according to  $\mu$ .
2. For a finite set  $I \subseteq \mathbb{N}$ , a random level- $I$  component, raw or rescaled, is chosen by first choosing  $n \in I$  uniformly, and independently choosing a raw or rescaled level- $n$  component, respectively.

*Notation 2.3.* When the symbols  $\mu^{x,i}$  and  $\mu_{x,i}$  appear inside an expression  $\mathbb{P}(\dots)$  or  $\mathbb{E}(\dots)$ , they will always denote random variables drawn according to the component distributions defined above. The range of  $i$  will be specified as needed.

---

<sup>5</sup>Definition 2.2 is motivated by Furstenberg's notion of a CP-distribution [9, 10, 14], which arise as limits as  $N \rightarrow \infty$  of the distribution of components of level  $1, \dots, N$ . These limits have a useful dynamical interpretation but in our finitary setting we do not require this technology.

The definition is best understood with some examples. For  $A \subseteq \mathcal{P}([0, 1]^d)$  we have

$$\begin{aligned}\mathbb{P}_{i=n}(\mu^{x,i} \in A) &= \int 1_A(\mu^{x,n}) d\mu(x) \\ \mathbb{P}_{0 \leq i \leq n}(\mu^{x,i} \in A) &= \frac{1}{n+1} \sum_{i=0}^n \int 1_A(\mu^{x,i}) d\mu(x).\end{aligned}$$

This notation implicitly defines  $x, i$  as random variables. Thus if  $A_0, A_1, \dots \subseteq \mathcal{P}([0, 1]^d)$  and  $D \subseteq [0, 1]^d$  we could write

$$\mathbb{P}_{0 \leq i \leq n}(\mu^{x,i} \in A_i \text{ and } x \in D) = \frac{1}{n+1} \sum_{i=0}^n \mu(x : \mu^{x,i} \in A_i \text{ and } x \in D).$$

Similarly, for  $f : \mathcal{P}([0, 1]^d) \rightarrow \mathbb{R}$  we have

$$\mathbb{E}_{n \leq i \leq n+k}(f(\mu^{x,i})) = \frac{1}{k+1} \sum_{i=n}^{n+k} \int f(\mu^{x,i}) d\mu(x).$$

When dealing with components of several measures  $\mu, \nu$ , we assume all choices of components  $\mu^{x,i}, \nu^{y,j}$  are independent unless otherwise stated. For instance,

$$\mathbb{P}_{i=n}(\mu^{x,i} \in A, \nu^{y,i} \in B) = \int \int 1_A(\mu^{x,n}) \cdot 1_B(\nu^{y,n}) d\mu(x) d\nu(y),$$

where as usual  $1_A$  is the indicator function on  $A$ , so  $1_A(\omega) = 1$  if  $\omega \in A$  and 0 otherwise. We use the same notation to average a real sequence, thus given  $a_n, \dots, a_{n+k} \in \mathbb{R}$ ,

$$\mathbb{E}_{n \leq i \leq n+k}(a_i) = \frac{1}{k+1} \sum_{i=n}^{n+k} a_i.$$

We record one obvious fact, which we will use repeatedly:

**Lemma 2.4.** For  $\mu \in \mathcal{P}(\mathbb{R}^d)$  and  $n \in \mathbb{N}$ ,

$$\mu = \mathbb{E}_{i=n}(\mu(\mathcal{D}_i(x)) \cdot \mu_{x,i}).$$

## 2.3 An inverse theorem

We first introduce finite-scale approximations of delta-masses and of uniform measures.

**Definition 2.5.** A measure  $\mu \in \mathcal{P}([0, 1])$  is  $\varepsilon$ -atomic if there is an interval  $I = B_\varepsilon(x)$  of length  $2\varepsilon$  such that  $\mu(\mathbb{R} \setminus I) < \varepsilon$ .

Note that if  $\mu \in \mathcal{P}([0, 1])$  is  $O(2^{-m})$ -atomic then  $H_m(\mu) = O(1/m)$ .

**Definition 2.6.** A measure  $\mu \in \mathcal{P}([0, 1])$  is  $(\varepsilon, m)$ -uniform if  $H_m(\mu) > 1 - \varepsilon$ .

Now, the approximate equality  $H_n(\mu * \nu) \approx H_n(\mu)$  occurs trivially if either  $\mu$  is close to uniform, or if  $\nu$  is close to atomic. As we saw in Section 2.1 there are other ways this can occur, but the following theorem shows that locally (for typical component measures) the two trivial scenarios are essentially the only ones.

**Theorem 2.7.** *For every  $\varepsilon > 0$  and integer  $m$  there is a  $\delta = \delta(\varepsilon, m) > 0$  such that for every  $n > n(\varepsilon, \delta, m)$ , the following holds: if  $\mu, \nu \in \mathcal{P}([0, 1])$  and*

$$H_n(\mu * \nu) < H_n(\mu) + \delta,$$

*then there are disjoint subsets  $I, J \subseteq \{1, \dots, n\}$  with  $|I \cup J| > (1 - \varepsilon)n$ , such that*

$$\begin{aligned} \mathbb{P}_{i=k}(\mu^{x,i} \text{ is } (\varepsilon, m)\text{-uniform}) &> 1 - \varepsilon && \text{for } k \in I \\ \mathbb{P}_{i=k}(\nu^{x,i} \text{ is } 2^{-m}\text{-atomic}) &> 1 - \varepsilon && \text{for } k \in J. \end{aligned}$$

The proof is given in Section 4.4. The proof is effective, but the dependencies we obtain between  $\delta$ ,  $m$ , and  $n$  are very bad and certainly far from optimal. We do not pursue this topic.

The alternatives in the theorem are not exclusive. To see this begin with a measure  $\mu \in [0, 1]$  such that  $\dim(\mu * \mu) = \dim \mu = 1/2$ , and such that  $\lim H_n(\mu) = \lim H_n(\mu * \mu) = \frac{1}{2}$  (such measures are not hard to construct by elementary means, or can be adapted from the more elaborate constructions in [19, 27]). By Marstrand's theorem, for a.e.  $t$  the scaled measure  $\nu(A) = \mu(tA)$  satisfies  $\dim \mu * \nu = 1$  and hence  $H_n(\mu * \nu) \rightarrow 1$ . But it is easy to verify that, as the conclusion of the theorem holds for the pair  $\mu, \mu$ , it holds for  $\mu, \nu$  as well.

Note that there is no assumption on the entropy of  $\nu$ , but if  $H_n(\nu)$  is sufficiently close to 0 the conclusion will automatically hold with  $I$  empty, and if  $H_n(\nu)$  is not too close to 0 then  $J$  cannot be too large relative to  $n$  (this follows from the comments after Definition 2.5 and Lemma 3.4 below). We obtain the following useful conclusion.

**Theorem 2.8.** *For every  $\varepsilon > 0$  and integer  $m$ , there is a  $\delta = \delta(\varepsilon, m) > 0$  such that for every  $n > n(\varepsilon, \delta, m)$  and every  $\mu \in \mathcal{P}([0, 1])$ , if*

$$\mathbb{P}_{0 \leq i \leq n}(H_m(\mu^{x,i}) < 1 - \varepsilon) > 1 - \varepsilon$$

*then for every  $\nu \in \mathcal{P}([0, 1])$*

$$H_n(\nu) > \varepsilon \quad \implies \quad H_n(\mu * \nu) \geq H_n(\mu) + \delta.$$

Specializing the above to self-convolutions we have the following result, which shows that constructions like the one described in Section 2.1 are, roughly, the only way that  $H_n(\mu * \mu) \approx H_n(\mu)$  can occur:

**Theorem 2.9.** *For every  $\varepsilon > 0$  and integer  $m$ , there is a  $\delta = \delta(\varepsilon, m) > 0$  such that for every sufficiently large  $n > n_*(\varepsilon, \delta, m)$  and every  $\mu \in \mathcal{P}([0, 1])$ , if*

$$H_n(\mu * \mu) < H_n(\mu) + \delta$$

*then there disjoint are subsets  $I, J \subseteq \{0, \dots, n\}$  with  $|I \cup J| \geq (1 - \varepsilon)n$  and such that*

$$\begin{aligned} \mathbb{P}_{i=k}(\mu^{x,i} \text{ is } (\varepsilon, m)\text{-uniform}) &> 1 - \varepsilon && \text{for } k \in I \\ \mathbb{P}_{i=k}(\mu^{x,i} \text{ is } 2^{-m}\text{-atomic}) &> 1 - \varepsilon && \text{for } k \in J. \end{aligned}$$

The theorems above hold more generally for compactly supported measures but the parameters will depend on the diameter of the support. It can also be extended to measures with unbounded support under additional assumptions, see Section 5.5.

### 3 Entropy, atomicity, uniformity

#### 3.1 Preliminaries on entropy

The Shannon entropy of a probability measure  $\mu$  with respect to a countable partition  $\mathcal{E}$  is given by

$$H(\mu, \mathcal{E}) = - \sum_{E \in \mathcal{E}} \mu(E) \log \mu(E),$$

where the logarithm is in base 2 and  $0 \log 0 = 0$ . The conditional entropy with respect to a countable partition  $\mathcal{F}$  is

$$H(\mu, \mathcal{E} | \mathcal{F}) = \sum_{F \in \mathcal{F}} \mu(F) \cdot H(\mu_F, \mathcal{E}),$$

where  $\mu_F = \frac{1}{\mu(F)} \mu|_F$  is the conditional measure on  $F$ . For a discrete probability measure  $\mu$  we write  $H(\mu)$  for the entropy with respect to the partition into points, and for a probability vector  $\alpha = (\alpha_1, \dots, \alpha_k)$  we write

$$H(\alpha) = - \sum \alpha_i \log \alpha_i.$$

We collect here some standard properties of entropy.

**Lemma 3.1.** *Let  $\mu, \nu$  be probability measures on a common space,  $\mathcal{E}, \mathcal{F}$  partitions of the underlying space and  $\alpha \in [0, 1]$ .*

1.  $H(\mu, \mathcal{E}) \geq 0$ , with equality if and only if  $\mu$  is supported on a single atom of  $\mathcal{E}$ .
2. If  $\mu$  is supported on  $k$  atoms of  $\mathcal{E}$  then  $H(\mu, \mathcal{E}) \leq k$ .
3. If  $\mathcal{F}$  refines  $\mathcal{E}$  (i.e.  $\forall F \in \mathcal{F} \exists E \in \mathcal{E} \text{ s.t. } F \subseteq E$ ) then  $H(\mu, \mathcal{F}) \geq H(\mu, \mathcal{E})$ .
4. If  $\mathcal{E} \vee \mathcal{F} = \{E \cap F : E \in \mathcal{E}, F \in \mathcal{F}\}$  denotes the join of  $\mathcal{E}$  and  $\mathcal{F}$ , then

$$H(\mu, \mathcal{E} \vee \mathcal{F}) = H(\mu, \mathcal{F}) + H(\mu, \mathcal{E} | \mathcal{F}).$$

5.  $H(\cdot, \mathcal{E})$  and  $H(\cdot, \mathcal{E} | \mathcal{F})$  are concave
6.  $H(\cdot, \mathcal{E})$  obeys the “convexity” bound

$$H\left(\sum \alpha_i \mu_i, \mathcal{E}\right) \leq \sum \alpha_i H(\mu_i, \mathcal{E}) + H(\alpha).$$

In particular, we note that for  $\mu \in \mathcal{P}([0, 1]^d)$  we have the bounds  $H(\mu, \mathcal{D}_m) \leq md$  and  $H(\mu, \mathcal{D}_{n+m} | \mathcal{D}_n) \leq md$ .

Although the function  $(\mu, m) \mapsto H(\mu, \mathcal{D}_m)$  is not weakly continuous, the following estimates provide usable substitutes.

**Lemma 3.2.** *Let  $\mu \in \mathcal{P}(\mathbb{R})$  and  $m \in \mathbb{N}$ .*

1. Given  $m \in \mathbb{N}$  and  $\mu \in \mathcal{P}(K)$  for some compact  $K \subseteq \mathbb{R}$ , there is a neighborhood  $U \subseteq \mathcal{P}(K)$  of  $\mu$  such that  $|H(\nu, \mathcal{D}_m) - H(\mu, \mathcal{D}_m)| \leq C_1$  for  $\nu \in U$ .
2. If  $\mathcal{E}, \mathcal{F}$  are partitions and each  $E \in \mathcal{F}$  intersects at most  $C_2$  elements of  $\mathcal{E}$  and vice versa, then  $|H(\mu, \mathcal{E}) - H(\mu, \mathcal{F})| < C_2 \log C_2$ .

3. If  $f, g : \mathbb{R}^d \rightarrow \mathbb{R}^k$  and  $\|f(x) - g(x)\| \leq C_3 2^{-m}$  for  $x \in \mathbb{R}^d$  then  $|H(f\mu, \mathcal{D}_m) - H(g\mu, \mathcal{D}_m)| \leq C'_3$  where  $C'_3$  depends only on  $k$ .
4. If  $\nu(\cdot) = \mu(\cdot + x_0)$  then  $|H(\mu, \mathcal{D}_m) - H(\nu, \mathcal{D}_m)| < C_4$ .
5. If  $C_5^{-1} \leq m'/m \leq C_5$ , then  $|H(\mu, \mathcal{D}_m) - H(\mu, \mathcal{D}_{m'})| \leq C'_5$ , where  $C'_5$  depends only on  $C_5$  and  $d$ .

Recall that the total variation distance between  $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$  is

$$\|\mu - \nu\| = \sup_A |\mu(A) - \nu(A)|,$$

where the supremum is over Borel sets  $A$ . This is a complete metric on  $\mathcal{P}(\mathbb{R}^d)$ .

**Lemma 3.3.** *For any bounded Borel set  $K \subseteq \mathbb{R}^d$  and any  $m \in \mathbb{N}$ , the function  $\mathcal{P}(K) \rightarrow \mathbb{R}$ ,  $\mu \mapsto H(\mu, \mathcal{D}_m)$ , is uniformly continuous in the total variation metric.*

### 3.2 Global entropy from local entropy

Recall from Section 2.2 the definition of the raw and re-scaled components  $\mu_{x,n}$ ,  $\mu^{x,n}$ , and note that

$$H(\mu^{x,n}, \mathcal{D}_m) = H(\mu_{x,n}, \mathcal{D}_{n+m}). \quad (17)$$

Also,

$$\begin{aligned} \mathbb{E}_{i=n} (H_m(\mu^{x,i})) &= \int \frac{1}{m} H(\mu^{x,n}, \mathcal{D}_m) d\mu(x) \\ &= \frac{1}{m} \int H(\mu_{x,n}, \mathcal{D}_{n+m}) d\mu(x) \\ &= \frac{1}{m} H(\mu, \mathcal{D}_{n+m} | \mathcal{D}_n). \end{aligned}$$

**Lemma 3.4.** *For  $r \geq 1$  and  $\mu \in \mathcal{P}([-r, r]^d)$  and integers  $m < n$ ,*

$$H_n(\mu) = \mathbb{E}_{0 \leq i \leq n} (H_m(\mu^{x,i})) + O\left(\frac{m}{n} + \frac{\log r}{n}\right).$$

*Proof.* The statement is equivalent to

$$H_n(\mu) = \frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{m} H(\mu, \mathcal{D}_{i+m} | \mathcal{D}_i) + O\left(\frac{m}{n} + \frac{\log r}{n}\right).$$

At the cost of adding  $O(m/n)$  to the error term we can delete up to  $m$  terms from the sum. Thus without loss of generality we may assume that  $n/m \in \mathbb{N}$ . When  $m = 1$ , iterating the conditional entropy formula gives

$$\sum_{i=0}^{n-1} H(\mu, \mathcal{D}_{i+1} | \mathcal{D}_i) = H(\mu, \mathcal{D}_n | \mathcal{D}_0) = H(\mu, \mathcal{D}_n) - O(\log r)$$

(since  $\mu \in \mathcal{P}([-r, r]^d)$  implies  $H(\mu, \mathcal{D}_0) = O(\log r)$ ), and the result follows on dividing by  $n$ . For general  $m$ , first decompose the sum according to the residue class of  $i \bmod m$

and apply the above to each one:

$$\begin{aligned} \sum_{i=0}^{n-1} \frac{1}{m} H(\mu, \mathcal{D}_{i+m} | \mathcal{D}_i) &= \frac{1}{m} \sum_{p=0}^{m-1} \left( \sum_{k=0}^{n/m-1} H(\mu, \mathcal{D}_{(k+1)m+p} | \mathcal{D}_{km+p}) \right) \\ &= \frac{1}{m} \sum_{p=0}^{m-1} H(\mu, \mathcal{D}_{n+p} | \mathcal{D}_p). \end{aligned}$$

Dividing by  $n$ , the result follows from the bound

$$\left| \frac{1}{n} H(\mu, \mathcal{D}_{n+p} | \mathcal{D}_p) - H_n(\mu) \right| < \frac{2m + \log r}{n},$$

which can be derived from the identities

$$\begin{aligned} H(\mu, \mathcal{D}_n) + H(\mu, \mathcal{D}_{n+p} | \mathcal{D}_n) &= H(\mu, \mathcal{D}_{n+p}) \\ &= H(\mu, \mathcal{D}_p) + H(\mu, \mathcal{D}_{n+p} | \mathcal{D}_p) \end{aligned}$$

together with the fact that  $H(\mu, \mathcal{D}_p) \leq p + \log r$  and  $H(\mu, \mathcal{D}_{rm+p} | \mathcal{D}_{rm}) \leq p$ , and recalling that  $0 \leq p < m$ .  $\square$

We have a similar lower bound for the entropy of a convolution in terms of convolutions of its components at each level.

**Lemma 3.5.** *Let  $r > 0$  and  $\mu, \nu \in \mathcal{P}([-r, r]^d)$ . Then for  $m < n \in \mathbb{N}$ ,*

$$H_n(\mu * \nu) \geq \mathbb{E}_{0 \leq i \leq n} (H_m(\mu^{x,i} * \nu^{y,i})) + O\left(\frac{1}{m} + \frac{m + \log r}{n}\right).$$

*Proof.* As in the previous proof, by introducing an error of  $O(m/n)$  we can assume that  $m$  divides  $n$ , and by the conditional entropy formula,

$$\begin{aligned} H(\mu * \nu, \mathcal{D}_n) &= \sum_{k=0}^{n/m-1} H(\mu * \nu, \mathcal{D}_{(k+1)m} | \mathcal{D}_{km}) + H(\mu * \nu, \mathcal{D}_0) \\ &= \sum_{k=0}^{n/m-1} H(\mu * \nu, \mathcal{D}_{(k+1)m} | \mathcal{D}_{km}) + O(\log r) \end{aligned}$$

since  $\mu * \nu$  is supported on  $[-2r, 2r]^d$ . Substituting the identity  $\mu * \nu = \mathbb{E}_{i=k}(\mu_{x,i} * \nu_{x,i})$ , and using concavity of entropy,

$$\begin{aligned} H(\mu * \nu, \mathcal{D}_n) &= \sum_{k=0}^{n/m-1} H(\mathbb{E}_{i=km}(\mu_{x,i} * \nu_{x,i}), \mathcal{D}_{(k+1)m} | \mathcal{D}_{km}) + O(\log r) \\ &\geq \sum_{k=0}^{n/m-1} \mathbb{E}_{i=km} (H(\mu_{x,i} * \nu_{x,i}, \mathcal{D}_{(k+1)m} | \mathcal{D}_{km})) + O(\log r) \\ &= \sum_{k=0}^{n/m-1} \mathbb{E}_{i=km} (H(\mu^{x,i} * \nu^{x,i}, \mathcal{D}_m | \mathcal{D}_0)) + O(\log r) \\ &= \sum_{k=0}^{n/m-1} \mathbb{E}_{i=km} (H(\mu^{x,i} * \nu^{x,i}, \mathcal{D}_m) + O(1)) + O(\log r) \\ &= \sum_{k=0}^{n/m-1} m \cdot \mathbb{E}_{i=km} (H_m(\mu^{x,i} * \nu^{x,i})) + O\left(\frac{n}{m} + \log r\right), \end{aligned}$$



where in the second-to-last equality we use the fact that  $\mu^{x,i} * \nu^{y,i}$  is supported on  $[0, 2)$  and therefore meets  $O(1)$  elements of  $\mathcal{D}_0$ . Dividing by  $n$ , we have shown that

$$H_n(\mu * \nu) \geq \frac{m}{n} \sum_{k=0}^{n/m-1} \mathbb{E}_{i=k} (H_m(\mu^{x,i} * \nu^{x,i})) + O\left(\frac{1}{m} + \frac{\log r}{n}\right).$$

Now do the same for the sum  $k = p$  to  $n/m + p$  for  $p = 0, 1, \dots, m-1$ . Averaging the resulting expressions gives the lemma.  $\square$

### 3.3 Atomicity and uniformity of components

The following technical results allow to pass from a measure to its component measures while preserving some of the concentration or uniformity properties of the original measure.

**Lemma 3.6.** *If  $\mu \in \mathcal{P}([0, 1])$  is  $\varepsilon$ -atomic and  $1 \leq m \leq \log(1/\varepsilon)$ , then*

$$\mathbb{P}_{i=m} (\mu^{x,i} \text{ is } 2^m \varepsilon\text{-atomic}) > 1 - \sqrt{2^{-m} \varepsilon}.$$

*Proof.* By definition  $1 - \varepsilon$  of the mass of  $\mu$  is concentrated on an interval  $W$  of length  $2\varepsilon$ . For  $D \in \mathcal{D}_m$  write  $T_D$  for the surjective homothety  $D \rightarrow [0, 1]^d$  and  $W^D = T_D W$ .

Take  $\delta = \sqrt{2^m \varepsilon}$  (note  $\delta \leq 1$ ) and let  $\mathcal{E} \subseteq \mathcal{D}_m$  denote the family of cells  $D$  such that

$$\mu_D(D \setminus W) = \mu^D([0, 1] \setminus (W^D)) \geq \delta.$$

It follows that

$$\varepsilon \geq \mu([0, 1] \setminus W) \geq \sum_{D \in \mathcal{E}} \mu(D \setminus W) = \sum_{D \in \mathcal{E}} \mu(D) \cdot \mu_D(D \setminus W) \geq \delta \cdot \mu(\cup \mathcal{E}),$$

so  $\mu(\cup \mathcal{E}) \leq \varepsilon/\delta = \sqrt{2^{-m} \varepsilon}$ . Hence  $\mu(\cup(\mathcal{D}_m \setminus \mathcal{E})) > 1 - \sqrt{2^{-m} \varepsilon}$ . Finally, for  $D \in \mathcal{D}_m \setminus \mathcal{E}$  we have  $\mu^D([0, 1] \setminus W^D) < \delta$  and  $W^D$  is an interval of length  $2^{m+1} \varepsilon$ , which implies that  $\mu^D$  is  $2^m \varepsilon$ -atomic, and the conclusion follows.  $\square$

**Lemma 3.7.** *If  $\mu \in \mathcal{P}([0, 1])$  is  $(\varepsilon, n)$ -uniform then for every  $1 \leq m < n$ ,*

$$\mathbb{P}_{0 \leq i \leq n} (\mu^{x,i} \text{ is } (\varepsilon', m)\text{-uniform}) > 1 - \varepsilon'$$

where  $\varepsilon' = \sqrt{\varepsilon} + O(\frac{m}{n})$ . In particular there is a subset  $I \subseteq \{0, \dots, n\}$  with  $|I| \geq (1 - \sqrt{\varepsilon'})n$  and

$$\mathbb{P}_{i=k} (\mu^{x,i} \text{ is } (\varepsilon', m)\text{-uniform}) > 1 - \sqrt{\varepsilon'} \quad \text{for } k \in I.$$

*Proof.* By Lemma 3.4 we have

$$\mathbb{E}_{0 \leq i \leq n} (H_m(\mu^{x,i})) = H_n(\mu) - O\left(\frac{m}{n}\right) > 1 - (\varepsilon + O(\frac{m}{n})) = 1 - (\varepsilon')^2,$$

so the first statement follows by Markov's inequality. Let  $I$  denote the set of  $0 \leq k \leq n$  such that  $\mathbb{P}_{i=k} (H_m(\mu^{x,i}) > 1 - \varepsilon') > 1 - \sqrt{\varepsilon'}$ . Since  $\mathbb{E}_{0 \leq i \leq n} (H_m(\mu^{x,i})) = \frac{1}{n+1} \sum_{k=0}^n \mathbb{E}_{i=k} (H_m(\mu^{x,i}))$ , by Markov's inequality again,  $|I| > (1 - \sqrt{\varepsilon'})n$ , as claimed.  $\square$

We also need a simple covering lemma.

**Lemma 3.8.** *Let  $I \subseteq \{1, \dots, n\}$  and  $\ell$  be given. Then there is a subset  $I' \subseteq I$  such that  $I \subseteq I' + [0, \ell]$  and  $(i + [0, \ell]) \cap (j + [0, \ell]) = \emptyset$  for distinct  $i, j \in I'$ .*

*Proof.* Define  $I'$  inductively, starting with the least element of  $I$  and at stage  $k$  adding the least element of  $I$  not covered by the sets  $j + [0, \ell]$  for  $j$  already in  $I'$ .  $\square$

## 4 Convolutions

### 4.1 The Berry-Esseen theorem and an entropy estimate

For  $\mu \in \mathcal{P}(\mathbb{R})$  let  $m(\mu)$  denote the mean, or barycenter, of  $\mu$ , given by

$$m(\mu) = \int x d\mu(x),$$

and let  $\text{Var}(\mu)$  denote its variance:

$$\text{Var}(\mu) = \int (x - m(\mu))^2 d\mu(x).$$

Recall that if  $\mu_1, \dots, \mu_k \in \mathcal{P}(\mathbb{R})$  then  $\mu = \mu_1 * \dots * \mu_k$  has mean  $m(\mu) = \sum_{i=1}^k m(\mu_i)$  and  $\text{Var}(\mu) = \sum_{i=1}^k \text{Var}(\mu_i)$ .

The Gaussian with mean  $m$  and variance  $\sigma^2$  is given by  $\gamma_{m,\sigma^2}(A) = \int_A \varphi((x - m)/\sigma) dx$ , where  $\varphi(x) = \sqrt{2\pi} \exp(-\frac{1}{2}|x|^2)$ . The central limit theorem asserts that, for  $\mu_1, \mu_2, \dots \in \mathcal{P}(\mathbb{R}^d)$  of positive variance, the convolutions  $\mu_1 * \dots * \mu_k$  can be re-scaled so that the resulting measure is close in the weak sense to a Gaussian measure. The Berry-Esseen inequalities quantify the rate of this convergence. We use the following variant from [7].

**Theorem 4.1.** *Let  $\mu_1, \dots, \mu_k$  be probability measures on  $\mathbb{R}$  with finite third moments  $\rho_i = \int |x|^3 d\mu_i(x)$ . Let  $\mu = \mu_1 * \dots * \mu_k$  and let  $\gamma$  be the Gaussian measure with the same mean and variance as  $\mu$ . Then for any interval  $I \subseteq \mathbb{R}$ ,*

$$|\mu(I) - \gamma(I)| \leq C_1 \cdot \frac{\sum_{i=1}^k \rho_i}{\text{Var}(\mu)^{3/2}},$$

where  $C_1 = C_1(d)$ . In particular, if  $\rho_i \leq C$  and  $\text{Var}(\mu_i) \geq c$  for constants  $c, C > 0$  then

$$|\mu(I) - \gamma(I)| = O_{c,C}(k^{-1/2}).$$

### 4.2 Multiscale analysis of repeated self-convolutions

In this section we show that for any measure  $\mu$ , every  $\delta > 0$ , every integer scale  $m \geq 2$ , and appropriately large  $k$ , the following holds: typical levels- $i$  components of the convolution  $\mu^{*k}$  are  $(\delta, m)$ -uniform, unless in  $\mu$  the level- $i$  components are typically  $2^{-m}$ -atomic. The main idea is to apply the Berry-Esseen theorem to convolutions of component measures.

**Proposition 4.2.** *Let  $\sigma > 0$ ,  $\delta > 0$ , and  $m \geq 2$  an integer. Then there exists an integer  $p = p_0(\sigma, \delta, m)$  such that for all  $k \geq k_0(\sigma, \delta, m)$ , the following holds:*

*Let  $\mu_1, \dots, \mu_k \in \mathcal{P}([0, 1])$ , let  $\mu = \mu_1 * \dots * \mu_k$ , and suppose that  $\text{Var}(\mu) \geq \sigma k$ . Then*

$$\mathbb{P}_{i=p-\lfloor \log \sqrt{k} \rfloor}(\mu^{x,i} \text{ is } (\delta, m)\text{-uniform}) > 1 - \delta. \quad (18)$$

*Proof.* It is a general fact that, for an absolutely continuous probability measure  $\gamma$ , for  $\gamma$ -a.e.  $x$ , as  $p \rightarrow \infty$  the components  $\gamma^{x,p}$  converge weak- $*$  to Lebesgue measure on  $[0, 1]$ , and in particular  $\mathbb{E}_{i=p}(H_m(\mu^{x,i})) \rightarrow 1$  as  $p \rightarrow \infty$ . In general this is a consequence of the martingale convergence theorem or the Lebesgue differentiation theorem, and there

is no guaranteed rate of convergence, but if  $\gamma$  has a continuous density function  $f$ , then convergence holds at every  $x$  for which  $f(x) > 0$ , and the rate depends only on  $f(x)$  and on the modulus of continuity of  $f$  at  $x$ . In particular for the family of Gaussians with mean 0 and variance in a given compact interval  $[\sigma^2, 1]$ , convergence is uniform in  $x$  and in the measure. Therefore, given  $\sigma, \delta > 0$  there is a  $p = p_0(\sigma, \delta, m)$  such that  $\mathbb{P}_{i=p}(H_m(\gamma^{x,i}) > 1 - \delta) > 1 - \delta$  for any Gaussian  $\gamma$  with  $\text{Var}(\gamma) \geq \sigma$ .

Now, if  $\mu_i$  and  $\mu$  are as in the statement and  $\mu'$  is  $\mu$  scaled by  $2^{-\lfloor \log \sqrt{k} \rfloor}$  (which is up to a constant factor the same as  $\sqrt{k}$ ), then by the Berry-Esseen theorem (Theorem 4.1)  $\mu'$  agrees with the Gaussian of the same mean and variance on intervals of length  $2^{-p-m}$  to a degree that can be made arbitrarily small by making  $k$  large in a manner depending on  $\sigma, p$ . In particular for large enough  $k$  this guarantees that  $\mathbb{P}_{i=p}(H_m((\mu')^{x,i}) > 1 - \delta) > 1 - \delta$ .

All that remains is to adjust the scale by a factor of  $2^{\lfloor \log \sqrt{k} \rfloor}$ . Then the same argument applied to  $\mu$  instead of the scaled  $\mu'$  gives  $\mathbb{P}_{i=p-\lfloor \log \sqrt{k} \rfloor}(H_m((\mu)^{x,i}) > 1 - \delta) > 1 - \delta$ , which is (18).  $\square$

We turn to repeated self-convolutions.

**Proposition 4.3.** *Let  $\sigma, \delta > 0$  and  $m \geq 2$  an integer. Then there exists  $p = p_1(\sigma, \delta, m)$  such that for sufficiently large  $k \geq k_1(\sigma, \delta, m)$ , the following holds.*

*Let  $\mu \in \mathcal{P}([0, 1])$ , fix an integer  $i_0 \geq 0$ , and write*

$$\lambda = \mathbb{E}_{i=i_0}(\text{Var}(\mu^{x,i})).$$

*If  $\lambda > \sigma$  then for  $j_0 = i_0 - \lfloor \log \sqrt{k} \rfloor + p$  and  $\nu = \mu^{*k}$  we have*

$$\mathbb{P}_{j=j_0}(\nu^{x,j} \text{ is } (\delta, m)\text{-uniform}) > 1 - \delta.$$

*Proof.* Fix  $\mu, \lambda$  and  $m$  be given. Fix  $p$  and  $k$  (we will later see how large they must be). Let  $i_0$  be as in the statement and  $j_0 = i_0 - \lfloor \log \sqrt{k} \rfloor + p$ .

Let  $\tilde{\mu}$  denote the  $k$ -fold self-product  $\tilde{\mu} = \mu \times \dots \times \mu$  and let  $\pi : (\mathbb{R})^k \rightarrow \mathbb{R}$  denote the addition map

$$\pi(x_1, \dots, x_k) = \sum_{i=1}^k x_i.$$

Then  $\nu = \pi\tilde{\mu}$ , and, since  $\tilde{\mu} = \mathbb{E}_{i=i_0}(\tilde{\mu}_{x,i})$ , we also have by linearity  $\nu = \mathbb{E}_{i=i_0}(\pi\tilde{\mu}_{x,i})$ . By concavity of entropy and an application of Markov's inequality, there is a  $\delta_1 > 0$ , depending only on  $\delta$ , such that the proposition will follow if we show that with probability  $> 1 - \delta_1$  over the choice of the component  $\tilde{\mu}_{x,i_0}$  of  $\tilde{\mu}$ , the measure  $\eta = \pi\tilde{\mu}_{x,i_0}$  satisfies

$$\mathbb{P}_{j=j_0}(\eta^{y,j} \text{ is } (\delta_1, m)\text{-uniform}) > 1 - \delta_1. \quad (19)$$

The random component  $\tilde{\mu}_{x,i_0}$  is itself a product measure  $\tilde{\mu}_{x,i} = \mu_{x_1,i_0} \times \dots \times \mu_{x_k,i_0}$ , and the marginal measures  $\mu_{x_j,i_0}$  of this product are distributed independently according to the distribution of the raw components of  $\mu$  at level  $i_0$ . Note that these components differ from the re-scaled components by a scaling factor of  $2^{i_0}$ , so the expected variance of the raw components is  $2^{-2i_0}\lambda$ . Recall that

$$\text{Var}(\pi(\mu_{x_1,i_0} \times \dots \times \mu_{x_k,i_0})) = \sum_{j=1}^k \text{Var}(\mu_{x_j,i_0}).$$

Thus for any  $\delta_2 > 0$ , by the weak law of large numbers, if  $k$  is large enough in a manner depending on  $\delta_2$  then with probability  $> 1 - \delta_2$  over the choice of  $\tilde{\mu}_{x,i_0}$  we will have<sup>6</sup>

$$\left| \frac{1}{k} \text{Var}(\pi \tilde{\mu}_{x,i_0}) - 2^{-2i_0} \lambda \right| < 2^{-2i_0} \delta_2. \quad (20)$$

We can choose  $\delta_2$  small in a manner depending on  $\sigma$ , so (20) implies

$$\text{Var}(\pi \tilde{\mu}_{x,i_0}) > 2^{-2i_0} \cdot k\sigma/2. \quad (21)$$

But now inequality (19) follows from an application of Proposition 4.2 with proper choice of parameters.  $\square$

**Lemma 4.4.** *Any  $\mu \in \mathcal{P}([0,1])$  is  $\text{Var}(\mu)^{1/3}$ -atomic. Conversely, if  $\mu$  is  $\varepsilon$ -atomic,  $\varepsilon < 1$  then  $\text{Var}(\mu) < \varepsilon$ .*

*Proof.* The second statement is trivial, the first is a simple consequence of Markov's inequality.  $\square$

**Theorem 4.5.** *Let  $\delta > 0$  and  $m \geq 2$  an integer. Then for  $k \geq k_2(\delta, m)$  and all sufficiently large  $n \geq n_2(\delta, m, k)$ , the following holds:*

*For any  $\mu \in \mathcal{P}([0,1])$  there are disjoint subsets  $I, J \subseteq \{1, \dots, n\}$  with  $|I \cup J| > (1 - \delta)n$  such that, writing  $\nu = \mu^{*k}$ ,*

$$\begin{aligned} \mathbb{P}_{i=q}(\nu^{x,i} \text{ is } (\delta, m)\text{-uniform}) &\geq 1 - \delta && \text{for } q \in I \\ \mathbb{P}_{i=q}(\mu^{x,i} \text{ is } 2^{-m}\text{-atomic}) &\geq 1 - \delta && \text{for } q \in J. \end{aligned}$$

*Proof.* Let  $\delta$  and  $m \geq 0$  be given, we may assume  $\delta < 1/2$ .

Let  $\tilde{\rho} : (0, 1] \rightarrow (0, 1]$  be a function such that  $\rho(\sigma)$  is small in a manner depending on  $\sigma, \delta, m$ . We shall specify the exact requirements in the course of the proof, and one may, if desired, collect them to give an explicit formula for  $\tilde{\rho}$ . We note that the definition of  $\tilde{\rho}$  uses the functions  $k_1(\cdot)$  and  $p_1(\cdot)$  from Proposition 4.3 and we assume, without loss of generality, that these functions are monotone in each of their arguments.

Our first condition on  $\tilde{\rho}$  will be that  $\tilde{\rho}(\sigma) < \sigma$ . Consider the decreasing sequence  $\sigma_0 > \sigma_1 > \dots$  defined by  $\sigma_0 = 1$  and  $\sigma_i = \tilde{\rho}(\sigma_{i-1})$ . Assume that  $k \geq k_1(\sigma_{[1+1/\delta^2]}, \delta, m)$ ; this expression can be taken for  $k_2(\delta, m)$ .

Fix  $\mu$  and  $n$  large, we shall later see how large an  $n$  is desirable. For  $0 \leq q \leq n$  write

$$\lambda_q = \mathbb{E}_{i=q}(\text{Var}(\mu^{x,i})).$$

Since the intervals  $(\sigma_i, \sigma_{i-1}]$  are disjoint, there is an integer  $1 \leq s \leq 1 + 1/\delta^2$  such that  $\mathbb{P}_{0 \leq q \leq n}(\lambda_q \in (\sigma_s, \sigma_{s-1}]) < \delta^2$ . For this  $s$  define

$$\begin{aligned} \sigma &= \sigma_{s-1} \\ \rho &= \tilde{\rho}(\sigma) = \sigma_s, \end{aligned}$$

and set

$$\begin{aligned} I' &= \{0 \leq q \leq n : \lambda_q > \sigma\} \\ J' &= \{0 \leq q \leq n : \lambda_q < \rho\}. \end{aligned}$$

---

<sup>6</sup>We use here the fact that we have a uniform bound for the rate of convergence in the weak law of large numbers for i.i.d. random variables  $X_1, X_2, \dots$ . In fact, the rate can be bounded in terms of the mean and variance of  $X_1$ . Here  $X_1$  is distributed like the variance  $\text{Var}(\mu_{x,i_0})$  of a random component of level  $i_0$ , and the mean and variance of  $X_1$  are bounded independently of  $\mu \in \mathcal{P}([0,1])$ .

Then by our choice of  $s$ ,

$$|I' \cup J'| > (1 - \delta^2)n.$$

We also write

$$p = p_1(\sigma, \delta, m)$$

and note that  $k \geq k_1(\sigma, \delta, m)$ .

Let  $\ell \geq 0$  be the integer

$$\ell = \lfloor \log \sqrt{k} \rfloor - p.$$

Since we may assume  $n$  large relative to  $\ell$ , by deleting at most  $\ell$  elements of  $I'$  we can assume that  $I' \subseteq [\ell, n]$  and that  $|I' \cup J'| > (1 - \delta^2)n$  still holds. Let

$$I = I' - \ell$$

By our choice of parameters and the previous proposition,

$$\mathbb{P}_{i=q}(\nu^{x,i} \text{ is } (\delta, m)\text{-uniform}) > 1 - \delta \quad \text{for } q \in I,$$

and also

$$\mathbb{E}_{i=q}(\text{Var}(\mu^{x,i})) = \lambda_q < \rho \quad \text{for } q \in J'.$$

By Markov's inequality,

$$\mathbb{P}_{i=q}(\text{Var}(\mu^{x,i}) < \sqrt{\rho}) > 1 - \sqrt{\rho} \quad \text{for } q \in J'.$$

By Lemma 4.4, this implies

$$\mathbb{P}_{i=q}(\mu^{x,i} \text{ is } \rho^{1/6}\text{-atomic}) > 1 - \sqrt{\rho} \quad \text{for } q \in J'.$$

Thus  $I, J'$  almost satisfy the conclusion of the theorem, except that they are not disjoint (even though  $I', J'$  were). To correct this, let

$$L = \lfloor \frac{1}{\delta^2}(\ell + 1) \rfloor.$$

By Lemma 3.6 and Lemma 3.8 and assuming as we may that  $n$  is sufficiently large compared to  $m$ , we can find  $J'' \subseteq J'$  such that  $J' \subseteq \bigcup_{q \in J''} [q, q + L]$ , the union is disjoint, and

$$\begin{aligned} \mathbb{P}_{i=t}(\mu^{x,i} \text{ is } \sqrt{2^L \rho^{1/6}}\text{-atomic}) &> (1 - \sqrt{\rho})(1 - \rho^{1/12}) \\ &\text{for } t \in \bigcup_{q \in J''} [q, q + L]. \end{aligned} \tag{22}$$

Let

$$J = \bigcup_{q \in J''} [q + \ell + 1, q + L].$$

The union is disjoint and by definition of  $L$  and  $J' \subseteq \bigcup_{q \in J''} [q, q + L]$ , its size is

$$|J| > (1 - \delta^2) \left| \bigcup_{q \in J''} [q, q + L] \right| \geq (1 - \delta^2)|J'|.$$

Assume now that  $\tilde{\rho}(\cdot)$  is such that  $\sqrt{2^L \rho^{1/6}} < 2^{-m}$  and  $(1 - \sqrt{\rho})(1 - \rho^{1/12}) < 1 - \delta^2$ . Then from the above we have

$$\mathbb{P}_{i=q}(\mu^{x,i} \text{ is } 2^{-m}\text{-atomic}) > (1 - \delta^2) \quad \text{for } q \in J.$$

We will be done if we show that  $I \cap J = \emptyset$ , since, using  $\delta < 1/2$  and  $|I' \cup J'| \geq (1 - \delta^2)n$  this implies

$$|I \cup J| = |I| + |J| \geq |I'| + (1 - \delta^2)|J'| \geq (1 - \delta^2)|I' \cup J'| \geq (1 - \delta^2)^2 n > (1 - \delta)n.$$

To see that  $I \cap J = \emptyset$ , suppose  $t \in I \cap J$ . Let

$$\pi = \mathbb{P}_{i=t+\ell} \left( \text{Var}(\mu^{x,i}) < \frac{\sigma}{2} \right)$$

Since  $t \in J$ , inequality (22) holds for  $t$ . By Lemma 3.6 and the relation between atomicity and variance, if  $\tilde{\rho}(\cdot)$  is suitably defined we will have  $\pi > (2 - 2\sigma)/(2 - \sigma)$ , hence

$$\lambda_{t+\ell} = \mathbb{E}_{i=t+\ell} (\text{Var}(\mu^{x,i})) < \frac{\sigma}{2}\pi + (1 - \pi) < \sigma$$

On the other hand  $t \in I$  means that  $t + \ell \in I'$  and we should have  $\lambda_{t+\ell} > \sigma$ . This contradiction shows that  $I \cap J = \emptyset$  and completes the proof.  $\square$

### 4.3 The Kaĭmanovich-Vershik-Tao lemma

The second ingredient in the proof of Theorem 2.7 is the following:

**Lemma 4.6** (Kaĭmanovich-Vershik, [15], Tao [31]). *Let  $\Gamma$  be a countable abelian group and let  $\mu, \nu \in \mathcal{P}(\Gamma)$  be probability measures with  $H(\mu) < \infty$ ,  $H(\nu) < \infty$ . Let*

$$\delta_k = H(\mu * (\nu^{*(k+1)})) - H(\mu * (\nu^{*k})).$$

*Then  $\delta_k$  is non-increasing in  $k$ . In particular,*

$$H(\mu * (\nu^{*k})) \leq H(\mu) + k \cdot (H(\mu * \nu) - H(\nu)).$$

This is the entropy analog of the Plünnecke-Rusza inequality in additive combinatorics, which states that if  $A, B \subseteq \mathbb{Z}$  are finite sets then  $|A + B|$  controls to some degree the growth of  $|A_0 + B^{\oplus k}|$ , where  $A_0 \subseteq A$  has size comparable to  $A$ . The result originates in a study by Kaĭmanovich and Vershik of random walks on groups and a version of it was recently rediscovered by Tao [31]. For completeness we give the proof.

*Proof.* Let  $X_0$  be a random variable distributed according to  $\mu$ , let  $Z_n$  be distributed according to  $\nu$ , and let all variables are independent. Set  $X_n = X_0 + Z_1 + \dots + Z_n$ , so the distribution of  $X_n$  is just  $\mu * \nu^{*n}$ . Furthermore, since  $G$  is abelian, given  $Z_1 = g$ , the distribution of  $X_n$  is the same as the distribution of  $X_{n-1} + g$  and hence  $H(X_n|Z_1) = H(X_{n-1})$ . We now compute:

$$\begin{aligned} H(Z_1|X_n) &= H(Z_1, X_n) - H(X_n) \\ &= H(Z_1) + H(X_n|Z_1) - H(X_n) \\ &= H(\nu) + H(\mu * \nu^{*(n-1)}) - H(\mu * \nu^{*n}). \end{aligned} \tag{23}$$

Since  $X_n$  is a Markov process, given  $X_n$ ,  $Z_1 = X_1 - X_0$  is independent of  $X_{n+1}$ , so

$$H(Z_1 | X_n) = H(Z_1 | X_n, X_{n+1}) \leq H(Z_1 | X_{n+1}).$$

Using (23) in both sides of the inequality above, we find that

$$H(\mu * \nu^{*(n-1)}) - H(\mu * \nu^{*n}) \leq H(\mu * \nu^{*n}) - H(\mu * \nu^{*(n+1)}),$$

which is the what we claimed.  $\square$

For the analogous statement for the scale- $n$  entropy of measures on  $\mathbb{R}$  we use a discretization argument. For  $m \in \mathbb{N}$  let

$$M_m = \left\{ \frac{k}{2^m} : k \in \mathbb{Z} \right\}$$

denote the group of  $2^m$ -adic rationals. Each  $D \in \mathcal{D}_m$  contains exactly one  $x \in M_m$ . Define the  $m$ -discretization map  $\sigma_m : \mathbb{R} \rightarrow M_m$  by  $\sigma_m(x) = v$  if  $\mathcal{D}_m(x) = \mathcal{D}_m(v)$ , so that  $\sigma_m(x) \in \mathcal{D}_m(x)$ .

We say that a measure  $\mu \in \mathcal{P}(\mathbb{R}^d)$  is  $m$ -discrete if it is supported on  $M_m$ , and for arbitrary  $\mu$  define its  $m$ -discretization to be its push-forward through the  $\sigma_m$ , explicitly:

$$\mu^{(m)} = \sum_{v \in M_m^d} \mu(\mathcal{D}_m(v)) \cdot \delta_v.$$

Clearly  $H_m(\mu) = H_m(\mu^{(m)})$ .

**Lemma 4.7.** *Given  $\mu_1, \dots, \mu_k \in \mathcal{P}(\mathbb{R})$  with  $H(\mu_i) < \infty$  and  $m \in \mathbb{N}$ ,*

$$|H_m(\mu_1 * \mu_2 * \dots * \mu_k) - H_m(\mu_1^{(m)} * \dots * \mu_k^{(m)})| = O(k/m).$$

*Proof.* Let  $\pi : \mathbb{R}^k \rightarrow \mathbb{R}$  denote the map  $(x_1, \dots, x_k) \mapsto \sum_{i=1}^k x_i$ . Then  $\mu_1 * \dots * \mu_k = \pi(\mu_1 \times \dots \times \mu_k)$  and  $\mu_1^{(m)} * \dots * \mu_k^{(m)} = \pi \circ \sigma_m(\mu_1 \times \dots \times \mu_k)$  (here we extend  $\sigma_m$  to  $(x_1, \dots, x_k) \mapsto (\sigma_m x_1, \dots, \sigma_m x_k)$ ). Now, it is easy to check that

$$|\pi(x_1, \dots, x_k) - \pi \circ \sigma_m(x_1, \dots, x_k)| = O(k)$$

so the desired entropy bound follows from Lemma 3.2 (3).  $\square$

**Proposition 4.8.** *Let  $\mu, \nu \in \mathcal{P}(\mathbb{R})$  with  $H_n(\mu), H_n(\nu) < \infty$ . Then*

$$H_n(\mu * (\nu^{*k})) \leq H_n(\mu) + k \cdot (H_n(\mu * \nu) - H_n(\mu)) + O\left(\frac{k}{n}\right). \quad (24)$$

*Proof.* Writing  $\tilde{\mu} = \mu^{(n)}$  and  $\tilde{\nu} = \nu^{(n)}$ , Theorem 4.6 implies

$$H(\tilde{\mu} * (\tilde{\nu}^{*k})) \leq H(\tilde{\mu}) + k \cdot (H(\tilde{\mu} * \tilde{\nu}) - H(\tilde{\nu})).$$

For  $n$ -discrete measures the entropy of the measure coincides with its entropy with respect to  $\mathcal{D}_n$ , so dividing this inequality by  $n$  gives (24) for  $\tilde{\mu}, \tilde{\nu}$  instead of  $\mu, \nu$ , and without the error term. The desired inequality follows from Lemma 4.7.  $\square$

We also will later need the following simple fact:

**Corollary 4.9.** *For  $m \in \mathbb{N}$  and  $\mu, \nu \in \mathcal{P}([-r, r]^d)$  with  $H_n(\mu), H_n(\nu) < \infty$ ,*

$$H_m(\mu * \nu) \geq H_m(\mu) - O\left(\frac{1}{m}\right).$$

*Proof.* This is immediate from the identity  $\mu * \nu = \int \mu * \delta_y d\nu(y)$ , concavity of entropy, and Lemma 3.2 (4) (note that  $\mu * \delta_y$  is a translate of  $\mu$ ).  $\square$

#### 4.4 Proof of the inverse theorem

For convenience we recall the formulation of Theorem 2.7:

**Theorem.** *For every  $\varepsilon > 0$  and  $m > m(\varepsilon)$ , there exists a  $\delta = \delta(\varepsilon, m)$  such that for all  $n > n(\varepsilon, m, \delta)$ , if  $\nu, \mu \in \mathcal{P}([0, 1])$  then either  $H_n(\mu * \nu) \leq H_n(\mu) + \delta$ , or there exist disjoint subsets  $I, J \subseteq \{0, \dots, n\}$  with  $|I \cup J| \geq (1 - \varepsilon)n$  and*

$$\begin{aligned} \mathbb{P}_{i=k}(\mu^{x,i} \text{ is } (\varepsilon, m)\text{-uniform}) &> 1 - \varepsilon && \text{for } k \in I \\ \mathbb{P}_{i=k}(\nu^{x,i} \text{ is } 2^{-m}\text{-atomic}) &> 1 - \varepsilon && \text{for } k \in J. \end{aligned}$$

*Proof.* Fix  $\varepsilon, m$  and choose  $k = k_2(\varepsilon, m)$  as in Theorem 4.5, with  $\delta = \varepsilon/2$ . We shall show that the conclusion holds if  $n$  is large relative to the previous parameters.

Let  $\mu, \nu \in \mathcal{P}([0, 1])$ . Denote

$$\tau = \nu^{*k}.$$

Assuming  $n$  is large enough, Theorem 4.5 provides us with disjoint subsets  $I, J \subseteq \{0, \dots, n\}$  with  $|I \cup J| > (1 - \varepsilon/2)n$  such that

$$\mathbb{P}_{i=k}(\tau^{x,i} \text{ is } (\frac{\varepsilon}{2}, m)\text{-uniform}) > 1 - \frac{\varepsilon}{2} \quad \text{for } k \in I \quad (25)$$

and

$$\mathbb{P}_{i=k}(\nu^{x,i} \text{ is } 2^{-m}\text{-atomic}) \geq 1 - \frac{\varepsilon}{2} \quad \text{for } k \in J. \quad (26)$$

Let  $I_0 \subseteq I$  denote the set of  $k$  such that

$$\mathbb{P}_{i=k}(\mu^{x,i} \text{ is } (\varepsilon, m)\text{-uniform}) > 1 - \varepsilon \quad \text{for } k \in I. \quad (27)$$

If  $|I_0| > (1 - \varepsilon)n$  we are done, since by (26) and (27), the pair  $I_0, J$  satisfy the second alternative of the theorem.

Otherwise, let  $I_1 = I \setminus I_0$ , so that  $|I_1| = |I| - |I_0| > \varepsilon n/2$ . We have

$$\mathbb{P}_{i=k}(\tau^{x,i} \text{ is } (\frac{\varepsilon}{2}, m)\text{-uniform and } \mu^{y,i} \text{ is not } (\varepsilon, m)\text{-uniform}) > \frac{\varepsilon}{2} \quad \text{for } k \in I_1.$$

For  $\mu^{x,i}, \tau^{y,i}$  in the event above, this just means that  $H_m(\tau^{y,i}) > H_m(\mu^{x,i}) + \varepsilon/2$  and hence  $H_m(\mu^{x,i} * \tau^{y,i}) \geq H_m(\mu^{x,i}) + \varepsilon/2 - O(1/m)$ . For any other pair  $\mu^{x,i}, \tau^{y,i}$  we have the trivial bound  $H_m(\mu^{x,i} * \tau^{y,i}) \geq H_m(\mu^{x,i}) - O(1/m)$ . Thus, using Lemmas 3.4, 3.5, 4.9,

$$\begin{aligned} H_n(\mu * \tau) &= \mathbb{E}_{0 \leq i \leq n}(H_m(\mu^{x,i} * \tau^{y,i})) + O(\frac{m}{n}) \\ &= \frac{|I_1|}{n+1} \mathbb{E}_{i \in I_1}(H_m(\mu^{x,i} * \tau^{y,i})) + \frac{n+1-|I_1|}{n+1} \mathbb{E}_{i \in I_1^c}(H_m(\mu^{x,i} * \tau^{y,i})) + O(\frac{1}{m} + \frac{m}{n}) \\ &> \frac{|I_1|}{n+1} \left( \mathbb{E}_{i \in I_1}(H_m(\mu^{x,i})) + (\frac{\varepsilon}{2})^2 \right) + \frac{n+1-|I_1|}{n+1} \mathbb{E}_{i \in I_1^c}(H_m(\mu^{x,i})) + O(\frac{1}{m} + \frac{m}{n}) \\ &= \mathbb{E}_{0 \leq i \leq n}(H_m(\mu^{x,i})) + (\frac{\varepsilon}{2})^3 + O(\frac{1}{m} + \frac{m}{n}) \\ &= H_n(\mu) + (\frac{\varepsilon}{2})^3 + O(\frac{1}{m} + \frac{m}{n}). \end{aligned}$$

So, assuming that  $\varepsilon$  was sufficiently small to begin with,  $m$  large with respect to  $\varepsilon$  and  $n$  large with respect to  $m$ , we have

$$H_n(\mu * \tau) > H_n(\mu) + \frac{\varepsilon^3}{10}.$$



On the other hand, by Proposition 4.8 above,

$$H_n(\mu * \tau) = H_n(\mu * \nu^{*k}) \leq H_n(\mu) + k \cdot (H_n(\mu * \nu) - H_n(\mu)) + O\left(\frac{k}{n}\right).$$

Assuming that  $n$  is large enough in a manner depending on  $\varepsilon$  and  $k$ , this and the previous inequality give

$$H_n(\mu * \nu) \geq H_n(\mu) + \frac{\varepsilon^3}{100k}.$$

This completes the proof of Theorem 2.7 for  $\delta = \varepsilon^3/100k$ .  $\square$

Theorems 2.8 and 2.9 are a formal consequences of Theorem 2.7, as discussed in Section 2.3.

## 5 Self-similar measures

### 5.1 Uniform entropy dimension and self-similar measures

The entropy dimension of a measure  $\theta \in \mathcal{P}(\mathbb{R})$  is the limit  $\lim_{n \rightarrow \infty} H_n(\theta)$ , assuming it exists; by Lemma 3.4, this is equivalent to  $\lim_{n \rightarrow \infty} \mathbb{E}_{0 \leq i \leq n} (H_m(\theta^{x,i})) = \alpha$  for all integers  $m$ . However, the convergence of the averages does not imply that the entropies of the components  $\theta^{x,i}$  concentrate around their mean, and examples show that they need not. We introduce the following stronger notion:

**Definition 5.1.** A measure  $\theta \in \mathcal{P}(\mathbb{R})$  has uniform entropy dimension  $\alpha$  if for every  $\varepsilon > 0$ , for large enough  $m$ ,

$$\liminf_{n \rightarrow \infty} \mathbb{P}_{0 \leq i \leq n} (|H_m(\theta^{x,i}) - \alpha| < \varepsilon) > 1 - \varepsilon. \quad (28)$$

Our main objective in this section is to prove:

**Proposition 5.2.** *Let  $\mu \in \mathcal{P}(\mathbb{R})$  be a self-similar measure and  $\alpha = \dim \mu$ . Then  $\mu$  has uniform entropy dimension  $\alpha$ .*

For simplicity we first consider the case that all the contractions in the IFS contract by the same ratio  $r$ . Thus, consider an IFS  $\Phi = \{\varphi_i\}_{i \in \Lambda}$  with  $\varphi_i(x) = r(x - a_i)$ ,  $0 < r < 1$ . We denote the attractor by  $X$  and without loss of generality assume that  $0 \in X \subseteq [0, 1]$ , which can always be arranged by a change of coordinates and may be seen not to affect the conclusions. Let  $\mu = \sum_{i \in \Lambda} p_i \cdot \varphi_i \mu$  be a self-similar measure and as usual write  $\varphi_i = \varphi_{i_1} \circ \dots \circ \varphi_{i_n}$  and  $p_i = p_{i_1} \cdot \dots \cdot p_{i_n}$  for  $i \in \Lambda^n$ .

Let

$$\alpha = \dim \mu$$

As we have noted already, self-similar measures have entropy dimension:

$$\lim_{n \rightarrow \infty} H_n(\mu) = \alpha \quad (29)$$

Fix  $\tilde{x} \in X$  and define probability measures

$$\mu_{x,k}^{[n]} = c \cdot \sum \{p_i \cdot \varphi_i \mu : i \in \Lambda^n, \varphi_i \tilde{x} \in \mathcal{D}_k(x)\},$$

where  $c = c(x, \tilde{x}, k, n)$  is a normalizing constant. Thus  $\mu_{x,k}^{[n]}$  differs from  $\mu_{x,k}$  in that, instead of restricting  $\mu = \sum_{i \in \Lambda^n} p_i \cdot \varphi_i \mu$  to  $\mathcal{D}_k(x)$ , we include or exclude each term

in its entirety depending on whether  $\varphi_i \tilde{x} \in \mathcal{D}_k(x)$ . Since  $\varphi_i \mu$  may not be supported entirely on either  $\mathcal{D}_k(x)$  or its complement, in general we have neither  $\mu_{x,k}^{[n]} \ll \mu_{x,k}$  nor  $\mu_{x,k} \ll \mu_{x,k}^{[n]}$ . Note that the definition of  $\mu_{x,k}^{[n]}$  depends on the point  $\tilde{x}$ , but this will not concern us.

For  $0 < \rho < 1$  it will be convenient to write

$$\ell(\rho) = \lceil \log \rho / \log r \rceil,$$

so  $\rho, r^{\ell(\rho)}$  differ by a multiplicative constant. Recall that  $\|\cdot\|$  denotes the total variation norm, see Section 3.1.

**Lemma 5.3.** *For every  $\varepsilon > 0$  there is a  $0 < \rho < 1$  such that, for all  $k$  and  $n = \ell(\rho 2^{-k})$ ,*

$$\mathbb{P}_{i=k} \left( \left\| \mu_{x,i} - \mu_{x,i}^{[n]} \right\| < \varepsilon \right) > 1 - \varepsilon. \quad (30)$$

Furthermore  $\rho$  can be chosen independently of  $\tilde{x}$  and of the coordinate system on  $\mathbb{R}$  (so the same bound holds for any translate of  $\mu$ ).

*Proof.* It is elementary that if  $\mu$  is atomic then it consists of a single atom. In this case the statement is trivial, so assume  $\mu$  is non-atomic. Then<sup>7</sup> given  $\varepsilon > 0$  there is a  $\delta > 0$  such that every interval of length  $\delta$  has  $\mu$ -mass  $< \varepsilon^2/2$ . Choose an integer  $q$  so that  $r^q < \delta/2$  and let  $\rho = r^q$ .

Let  $k \in \mathbb{N}$  and  $\ell = \ell(2^{-k})$ , so that  $2^{-k} \cdot r \leq r^\ell \leq 2^{-k}$ . Let  $i \in \Lambda^\ell$  and consider those  $j \in \Lambda^q$  such that  $\varphi_{ij} \mu$  is not supported on an element of  $\mathcal{D}_k$ . Then  $\varphi_{ij} \mu$  is supported on the interval  $J$  of length  $\delta$  centered at one of the endpoints of an element of  $\mathcal{D}_k$ . Since  $\varphi_i \mu$  can give positive mass to at most two such intervals  $J$ , and  $\varphi_i \mu(J) < \varepsilon^2/2$  for each such  $J$ , we conclude that in the representation  $\mu_i = \frac{1}{p_i} \sum_{j \in \Lambda^q} p_{ij} \cdot (\varphi_{ij} \mu)$ , at least  $1 - \varepsilon^2$  of the mass comes from terms that are supported entirely on just one element of  $\mathcal{D}_k$ . Therefore the same is true in the representation  $\mu = \sum_{u \in \Lambda^{\ell+q}} p_u \cdot \varphi_u \mu$ . The inequality (30) now follows by an application of the Markov inequality. Finally, Since our choice of parameters did not depend on  $\tilde{x}$  and is invariant under translation of  $\mu$  and of the IFS, the last statement holds.  $\square$

**Lemma 5.4.** *For  $\varepsilon > 0$ , for large enough  $m$  and all  $k$ ,*

$$\mathbb{P}_{i=k} \left( H_m(\mu^{x,i}) > \alpha - \varepsilon \right) > 1 - \varepsilon,$$

and the same holds for any translate of  $\mu$ .

*Proof.* Let  $\varepsilon > 0$  be given. Choose  $0 < \varepsilon' < \varepsilon$  sufficiently small that  $\|\nu - \nu'\| < \varepsilon'$  implies  $|H(\nu, \mathcal{D}_m) - H(\nu', \mathcal{D}_m)| < \varepsilon/2$  for every  $\nu, \nu' \in \mathcal{P}([0, 1]^d)$  (Lemma 3.3). Let  $\rho$  be as in the previous lemma chosen with respect to  $\varepsilon'$ . Assume that  $m$  is large enough that  $|H_m(\mu') - \alpha| < \varepsilon/2$  whenever  $\mu'$  is  $\mu$  scaled by a factor of at most  $\rho$  ( $m$  exists by (29) and Lemma 3.2 (5)). Now fix  $k$  and let  $\ell = \ell(\rho 2^{-k})$ . By the previous lemma and choice of  $\varepsilon'$ , it is enough to show that  $\frac{1}{m} H(\mu_{x,k}^{[\ell]}, \mathcal{D}_{k+m}) > \alpha - \varepsilon/2$ . But this follows from the fact that  $\mu_{x,k}^{[\ell]}$  is a convex combination of measures  $\mu_j$  for  $j \in \Lambda^\ell$ , our choice of  $m$  and  $\ell$ , and concavity of entropy.  $\square$

<sup>7</sup>This is the only part of the proof of Theorem 1.3 which is not effective, but with a little more work one could make it effective in the sense that, if  $\liminf -\log \Delta^{(n)} = M < \infty$ , then at arbitrarily small scales one can obtain estimates of the continuity of  $\mu$  in terms of  $M$ .

We now prove Proposition 5.2. Let  $0 < \varepsilon < 1$  be given and fix an auxiliary parameter  $\varepsilon' < \varepsilon/2$ . We first show that this holds for  $m$  large in a manner depending on  $\varepsilon$ . Specifically let  $m$  be large enough that the previous lemma applies for the parameter  $\varepsilon'$ . In particular for any  $n$ ,

$$\mathbb{P}_{0 \leq i \leq n} (H_m(\mu^{x,i}) > \alpha - \varepsilon') > 1 - \varepsilon'. \quad (31)$$

By (29), for  $n$  large enough we have  $|H_n(\mu) - \alpha| < \varepsilon'/2$ , so by Lemma 3.4, for large enough  $n$  we have

$$|\mathbb{E}_{0 \leq i \leq n} (H_m(\mu^{x,i})) - \alpha| < \varepsilon'.$$

Since  $H_m(\mu^{x,i}) \geq 0$ , the last two equalities imply

$$\mathbb{P}_{0 \leq i \leq n} (H_m(\mu^{x,i}) < \alpha + \varepsilon'') > 1 - \varepsilon''$$

for some  $\varepsilon''$  that tend to 0 with  $\varepsilon'$ . Thus, choosing  $\varepsilon'$  small enough, the last inequality and (31) give (28), as desired.

When the contraction ratios are not uniform,  $\varphi_i = r_i x + a_i$ , some minor changes are needed in the proof. Given  $n$ , let  $\Lambda^{(n)}$  denote the set of  $i \in \Lambda^* = \bigcup_{m=1}^{\infty} \Lambda^m$  such that  $r_i < r^n \leq r_j$ , where  $j$  is the same as  $i$  but with the last symbol deleted (so its length is one less than  $i$ ). This ensures that  $\{r_i\}_{i \in \Lambda^{(n)}}$  are all within a multiplicative constant of each other (this constant is  $\min\{r_j : j \in \Lambda\}$ ). It is easy to check that  $\Lambda^{(n)}$  is a section of  $\Lambda^*$  in the sense that every sequence  $i \in \Lambda^*$  with  $r_i < r^n$  has a unique prefix in  $\Lambda^{(n)}$ . Now define  $\mu_{x,k}^{[n]}$  as before, but using  $\varphi_i \mu$  for  $i \in \Lambda^{(n)}$ , i.e.

$$\mu_{x,k}^{[n]} = c \cdot \sum \left\{ p_i \cdot \varphi_i \mu : i \in \Lambda^{(n)}, \varphi_i \tilde{x} \in \mathcal{D}_k(x) \right\}.$$

With this modification all the previous arguments now go through.

Finally, let us note the following consequence of the inverse theorem (Theorem 2.8).

**Corollary 5.5.** *For every  $0 < \alpha < 1$  and  $\varepsilon > 0$  there is a  $\delta > 0$  and such that the following holds: If  $\mu \in \mathcal{P}([0, 1])$  has uniform entropy  $\alpha$ , then for all large enough  $n$  and every  $\nu \in \mathcal{P}([0, 1])$ ,*

$$H_n(\nu) > \varepsilon \quad \implies \quad H_n(\mu * \nu) \geq H_n(\mu) + \delta.$$

Similar conclusions hold for dimension.

## 5.2 Proof of Theorem 1.3

We again begin with the uniformly contracting case,  $\varphi_i = rx + a_i$ , and continue with the notation from the previous section, in particular assume that 0 is in the attractor. Recall from the introduction that

$$\nu^{(n)} = \sum_{i \in \Lambda^n} p_i \cdot \delta_{\varphi_i(0)}.$$

Define

$$\tau^{(n)}(A) = \mu(r^{-n}A).$$

One may verify easily, using the assumption  $0 \in X$ , that

$$\mu = \nu^{(n)} * \tau^{(n)}. \quad (32)$$

As in the introduction, write

$$n' = \lceil n \log(1/r) \rceil.$$

Thus  $\tau^{(n)}$  is  $\mu$  scaled down by a factor of  $r^n = 2^{-n'}$  and translated. Using (29), Lemma 3.2, and the fact that  $\tau^{(n)}$  is supported on an interval of order  $r^n = 2^{-n'}$ , we have

$$\lim_{n \rightarrow \infty} \frac{1}{n'} H(\nu^{(n)}, \mathcal{D}_{n'}) = \lim_{n \rightarrow \infty} \frac{1}{n'} H(\mu, \mathcal{D}_{n'}) = \dim \mu = \alpha.$$

Suppose now that  $\alpha < 1$ . Fix a large  $q$  and consider the identity

$$\begin{aligned} \frac{1}{qn} H(\mu, \mathcal{D}_{qn}) &= \frac{n'}{qn} \cdot \left( \frac{1}{n'} H(\mu, \mathcal{D}_{n'}) \right) + \frac{qn - n'}{qn} \cdot \left( \frac{1}{qn - n'} H(\mu, \mathcal{D}_{qn} | \mathcal{D}_{n'}) \right) \\ &= \frac{\lceil \log(1/r) \rceil}{q} \left( \frac{1}{n'} H(\mu, \mathcal{D}_{n'}) \right) + \frac{q - \lceil \log(1/r) \rceil}{q} \left( \frac{1}{qn - n'} H(\mu, \mathcal{D}_{qn} | \mathcal{D}_{n'}) \right). \end{aligned}$$

The left hand side and the term  $\frac{1}{n'} H(\mu, \mathcal{D}_{n'})$  on the right hand side both tend to  $\alpha$  as  $n \rightarrow \infty$ . Since  $r, q$  are independent of  $n$  we conclude that

$$\lim_{n \rightarrow \infty} \frac{1}{qn - n'} H(\mu, \mathcal{D}_{qn} | \mathcal{D}_{n'}) = \alpha. \quad (33)$$

From the identity  $\nu^{(n)} = \mathbb{E}_{i=n'}(\nu_{i,x}^{(n)})$  and linearity of convolution,

$$\mu = \nu^{(n)} * \tau^{(n)} = \mathbb{E}_{i=n'} \left( \nu_{y,i}^{(n)} * \tau^{(n)} \right).$$

Also, each measure  $\nu_{y,i}^{(n)} * \tau^{(n)}$  is supported on an interval of length  $O(2^{-n'})$  so

$$|H(\nu_{i,y}^{(n)} * \tau^{(n)}, \mathcal{D}_{qn} | \mathcal{D}_{n'}) - H(\nu_{i,y}^{(n)} * \tau^{(n)}, \mathcal{D}_{qn})| = O(1).$$

By concavity of conditional entropy (Lemma 3.1 (5)),

$$\begin{aligned} H(\mu, \mathcal{D}_{qn} | \mathcal{D}_{n'}) &= H(\nu^{(n)} * \tau^{(n)}, \mathcal{D}_{qn} | \mathcal{D}_{n'}) \\ &\geq \mathbb{E}_{i=n'} \left( H(\nu_{y,i}^{(n)} * \tau^{(n)}, \mathcal{D}_{qn} | \mathcal{D}_{n'}) \right) \\ &= \mathbb{E}_{i=n'} \left( H(\nu_{y,i}^{(n)} * \tau^{(n)}, \mathcal{D}_{qn}) \right) + O(1), \end{aligned}$$

so by (33),

$$\limsup_{n \rightarrow \infty} \frac{1}{qn - n'} \mathbb{E}_{i=n'} \left( H(\nu_{y,i}^{(n)} * \tau^{(n)}, \mathcal{D}_{qn}) \right) \leq \alpha. \quad (34)$$

Now, we also know that

$$\lim_{n \rightarrow \infty} \frac{1}{qn - n'} H(\tau^{(n)}, \mathcal{D}_{qn}) = \alpha, \quad (35)$$

since, up to a re-scaling, this is just (29) (we again used the fact that  $\tau^{(n)}$  is supported on intervals of length  $2^{-n'}$ ). By Lemma 4.9, for every component  $\nu_{y,i}^{(n)}$ ,

$$\frac{1}{qn - n'} H(\nu_{y,i}^{(n)} * \tau^{(n)}, \mathcal{D}_{qn}) \geq \frac{1}{qn - n'} H(\tau^{(n)}, \mathcal{D}_{qn}) + O\left(\frac{1}{qn - n'}\right).$$

Therefore for every  $\delta > 0$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P}_{i=n'} \left( \frac{1}{qn - n'} H(\nu_{y,i}^{(n)} * \tau^{(n)}, \mathcal{D}_{qn}) > \alpha - \delta \right) = 1$$

which, combined with (34), implies that for every  $\delta > 0$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P}_{i=n'} \left( \left| \frac{1}{qn - n'} H(\nu_{y,i}^{(n)} * \tau^{(n)}, \mathcal{D}_{qn}) - \alpha \right| < \delta \right) = 1,$$

and replacing  $\alpha$  with the limit in (40), we have that for all  $\delta > 0$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P}_{i=n'} \left( \left| \frac{1}{qn - n'} H(\nu_{y,i}^{(n)} * \tau^{(n)}, \mathcal{D}_{qn}) - \frac{1}{qn - n'} H(\tau^{(n)}, \mathcal{D}_{qn}) \right| < \delta \right) = 1. \quad (36)$$

Now let  $\varepsilon > 0$ . By Proposition 5.2 and the assumption that  $\alpha < 1$ , for small enough  $\varepsilon$ , large enough  $m$  and all sufficiently large  $n$ ,

$$\begin{aligned} \mathbb{P}_{n' < i \leq qn'} \left( H_m((\tau^{(n)})^{x,i}) < 1 - \varepsilon \right) &\geq \mathbb{P}_{n' < i \leq qn'} \left( H_m((\tau^{(n)})^{x,i}) < \alpha + \varepsilon \right). \\ &> 1 - \varepsilon \end{aligned}$$

Choose  $\delta > 0$  smaller than the constant of the same name in the conclusion of Theorem 2.8. Then, for sufficiently large  $n$ , we can apply Theorem 2.8 to the components  $\nu_{y,i}^{(n)}$  in the event in equation (36) (for this we re-scale by  $2^{n'}$  and note that the measures  $\nu_{y,n'}^{(n)}$  are supported on level- $n'$  dyadic cells and  $\tau^{(n)}$  is supported on an interval of the same order of magnitude). We conclude that every component  $\nu_{y,i}^{(n)}$  in the event in question satisfies  $\frac{1}{qn - n'} H(\nu_{y,i}^{(n)}, \mathcal{D}_{qn}) < \varepsilon$ , and hence by (36),

$$\lim_{n \rightarrow \infty} \mathbb{P}_{i=n'} \left( \frac{1}{qn - n'} H(\nu_{y,i}^{(n)}, \mathcal{D}_{qn}) < \varepsilon \right) = 1.$$

Thus, from the definition of conditional entropy and the last equation,

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{qn - n'} H(\nu^{(n)}, \mathcal{D}_{qn} | \mathcal{D}_{n'}) &= \lim_{n \rightarrow \infty} \frac{1}{qn - n'} \mathbb{E}_{i=n'} \left( H(\nu_{y,i}^{(n)}, \mathcal{D}_{qn}) \right) \\ &= \lim_{n \rightarrow \infty} \mathbb{E}_{i=n'} \left( \frac{1}{qn - n'} H(\nu_{y,i}^{(n)}, \mathcal{D}_{qn}) \right) \\ &< \varepsilon. \end{aligned}$$

Since  $\varepsilon$  was arbitrary, this is Theorem 1.3.

### 5.3 Proof of Theorem 1.4 (the non-uniformly contracting case)

We now consider the situation for general IFS, in which the contraction  $r_i$  of  $\varphi_i$  is not constant. Again assume that 0 is in the attractor. Let  $r = \prod_{i \in \Lambda} r_i^{p_i}$ ,  $n' = \log_2(1/r)$  as in the introduction, and define  $\nu^{(n)}, \tilde{\nu}^{(n)}$  as before. Given  $n$ , let

$$R_n = \{r_i : i \in \Lambda^n\}.$$

Note that  $|R_n| = O(n^{|\Lambda|})$ . Therefore  $H(\tilde{\nu}^{(n)}, \{\mathbb{R}\} \times \mathcal{F}) = O(\log n)$ , and consequently for all  $k$

$$H(\tilde{\nu}^{(n)}, \tilde{\mathcal{D}}_k) = H(\nu^{(n)}, \mathcal{D}_k) + O(\log n).$$

Thus

$$H(\tilde{\nu}^{(n)}, \tilde{\mathcal{D}}_{qn} | \tilde{\mathcal{D}}_{n'}) = H(\nu^{(n)}, \mathcal{D}_{qn} | \mathcal{D}_n) + O(\log n),$$

and our goal reduces to proving that for every  $q > 1$ ,

$$\frac{1}{qn} H(\nu^{(n)}, \mathcal{D}_{qn} | \mathcal{D}_{n'}) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Furthermore, for every  $\varepsilon > 0$

$$H(\nu^{(n)}, \mathcal{D}_{qn} | \mathcal{D}_{(1-\varepsilon)n'}) = H(\nu^{(n)}, \mathcal{D}_{qn} | \mathcal{D}_{n'}) - O(\varepsilon n),$$

so it will suffice for us to prove that

$$\limsup_{n \rightarrow \infty} \frac{1}{qn} H(\nu^{(n)}, \mathcal{D}_{qn} | \mathcal{D}_{(1-\varepsilon)n}) = o(1) \quad \text{as } \varepsilon \rightarrow 0.$$

Fix  $\varepsilon > 0$ . For  $t \in R_n$  let

$$\begin{aligned} \Lambda^{n,t} &= \{i \in \Lambda^n : r_i = t\} \\ p^{n,t} &= \sum_{i \in \Lambda^{n,t}} p_i, \end{aligned}$$

so  $\{p^{n,t}\}_{t \in R_n}$  is a probability vector. It will sometimes be convenient to consider  $i \in \Lambda^n$ ,  $i \in \Lambda^{n,t}$  and  $t \in R_n$  as random elements drawn according to the probabilities  $p_i$ ,  $p_i/p^{n,t}$ , and  $p^{n,t}$ , respectively. Then we interpret expressions such as  $\mathbb{P}_{i \in \Lambda^n}(A)$ ,  $\mathbb{P}_{i \in \Lambda^{n,t}}(A)$  and  $\mathbb{P}_{t \in R_n}(A)$  in the obvious manner, and similarly expectations. With this notation, we can define

$$\nu^{(n,t)} = \mathbb{E}_{i \in \Lambda^{n,t}}(\delta_{\varphi_i(0)}) = \frac{1}{p^{n,t}} \sum_{i \in \Lambda^{n,t}} p_i \cdot \delta_{\varphi_i(0)}.$$

This a probability measure on  $\mathbb{R}$  representing the part of  $\nu^{(n)}$  coming from contractions by  $t$ ; indeed,

$$\nu^{(n)} = \mathbb{E}_{t \in R_n}(\nu^{(n,t)}). \quad (37)$$

For  $t > 0$  let  $\tau^{(t)}$  be the measure

$$\tau^{(t)}(A) = \tau(tA)$$

(note that we are no longer using logarithmic scale, so the measure that was previously denoted  $\tau^{(n)}$  is now  $\tau^{(2^{-n})}$ ). We then have

$$\mu = \mathbb{E}_{t \in R_n}(\nu^{(n,t)} * \tau^{(t)}). \quad (38)$$

Fix  $\varepsilon > 0$ . Arguing as in the previous section, using equation (38) and concavity of entropy, we have

$$\begin{aligned} \alpha &= \lim_{n \rightarrow \infty} \frac{1}{qn - (1-\varepsilon)n'} H(\mu, \mathcal{D}_{qn} | \mathcal{D}_{(1-\varepsilon)n'}) \\ &\geq \limsup_{n \rightarrow \infty} \frac{1}{qn - (1-\varepsilon)n'} \mathbb{E}_{t \in R_n} \left( H(\nu^{(n,t)} * \tau^{(t)}, \mathcal{D}_{qn} | \mathcal{D}_{(1-\varepsilon)n'}) \right). \end{aligned} \quad (39)$$

By the law of large numbers,

$$\lim_{n \rightarrow \infty} \mathbb{P}_{i \in \Lambda^n} \left( 2^{-(1+\varepsilon)n'} < r_i < 2^{-(1-\varepsilon)n'} \right) = 1,$$

or, equivalently,

$$\lim_{n \rightarrow \infty} \mathbb{P}_{t \in R_n} \left( 2^{-(1+\varepsilon)n'} < t < 2^{-(1-\varepsilon)n'} \right) = 1. \quad (40)$$

Using  $H_k(\mu) \rightarrow \alpha$  and the definition of  $\tau^{(t)}$ , we conclude that

$$\lim_{n \rightarrow \infty} \mathbb{P}_{t \in R_n} \left( \frac{1}{qn - (1-\varepsilon)n'} H(\tau^{(t)}, \mathcal{D}_{qn}) \geq (1-\varepsilon)\alpha \right) = 1.$$

Also, since  $\tau^{(t)}$  is supported on an interval of order  $t$ , from (40), (39) and concavity of entropy,

$$\begin{aligned} \alpha &\geq \limsup_{n \rightarrow \infty} \frac{1}{qn - (1-\varepsilon)n'} \mathbb{E}_{t \in R_n} \mathbb{E}_{i=n'} \left( H(\nu_{y,i}^{(n,t)} * \tau^{(t)}, \mathcal{D}_{qn} | \mathcal{D}_{(1-\varepsilon)n'}) \right) \\ &= \limsup_{n \rightarrow \infty} \frac{1}{qn - (1-\varepsilon)n'} \mathbb{E}_{t \in R_n} \mathbb{E}_{i=n'} \left( H(\nu_{y,i}^{(n,t)} * \tau^{(t)}, \mathcal{D}_{qn}) \right). \end{aligned} \quad (41)$$

This is the analogue of Equation 34 in the proof of the uniformly contracting case and from here one proceeds exactly as in that proof to conclude that there is a function  $\delta(\varepsilon)$ , tending to 0 as  $\varepsilon \rightarrow 0$ , such that

$$\mathbb{P}_{t \in R_n} \left( \mathbb{P}_{i=n'} \left( \frac{1}{qn - (1-\varepsilon)n} H(\nu^{(n,t)}, \mathcal{D}_{qn}) < \delta(\varepsilon) \right) \right) = 1.$$

Now, using Equation (37) and the fact that the entropy of the distribution  $\{p^{(n,t)}\}_{t \in R_n}$  is  $o(n)$  as  $n \rightarrow \infty$ , by Lemma 3.1 (6) one concludes that

$$\limsup_{n \rightarrow \infty} H(\nu^{(n)}, \mathcal{D}_{qn} | \mathcal{D}_{(1-\varepsilon)n'}) \leq \delta(\varepsilon),$$

which is what we wanted to prove.

## 5.4 Transversality and the dimension of exceptions

In this section we prove Theorem 1.8. Let  $I \subseteq \mathbb{R}$  be a compact interval for  $t \in I$  and let  $\Phi_t = \{\varphi_{i,t}\}_{i \in \Lambda}$  be an IFS,  $\varphi_{i,t}(x) = r_i(t)(x - a_i(t))$ . We define  $\varphi_{i,t}$  and  $r_i(t)$  for  $i \in \Lambda^n$  as usual, set  $\Delta_{i,j}(t) = \varphi_{i,t}(0) - \varphi_{j,t}(0)$  when  $i, j \in \Lambda^n$  and for  $i, j \in \Lambda^\mathbb{N}$  define  $\Delta_{i,j}(t) = \lim \Delta_{i_1 \dots i_n, j_1 \dots j_n}(t)$  (this is well defined since  $\lim \varphi_{i_1 \dots i_n}(0)$  converges, in fact exponentially, as  $n \rightarrow \infty$ ).

For  $i, j \in \Lambda^n$  or  $i, j \in \Lambda^\mathbb{N}$  let  $i \wedge j$  denote the longest common initial segment of  $i, j$ , and  $|i \wedge j|$  its length, so  $|i \wedge j| = \min\{k : i_k \neq j_k\} - 1$ . Let

$$r_{\min} = \min_{i \in \Lambda} \min_{t \in I} |r_i(t)|,$$

so  $0 < r_{\min} < 1$ . For a  $C^k$ -function  $F : I \rightarrow \mathbb{R}$  write  $F^{(p)} = \frac{d^p}{dt^p} F$ , and

$$\|F\|_{I,k} = \max_{p \in \{0, \dots, k\}} \max_{t \in I} |F^{(p)}(t)|.$$

In particular we write

$$R_k = \max_{i \in \Lambda} \|r_i\|_{I,k}.$$

**Definition 5.6.** The family  $\{\Phi_t\}_{t \in I}$  is *transverse of order  $k$*  if  $r_i(\cdot), a_i(\cdot)$  are  $k$ -times continuously differentiable and there is a constant  $c > 0$  such that for every  $n \in \mathbb{N}$  and distinct  $i, j \in \Lambda^n$ ,

$$\forall t_0 \in I \quad \exists p \in \{0, 1, 2, \dots, k\} \quad \text{such that} \quad |\Delta_{i,j}^{(p)}(t_0)| \geq c \cdot |i \wedge j|^{-p} \cdot r_{i \wedge j}(t_0). \quad (42)$$

The classical notion of transversality roughly corresponds to the case  $k = 1$  in this definition, see e.g. [21, Definition 2.7]. Unlike the classical notion, which either fails or is difficult to verify in many cases of interest, higher-order transversality holds almost automatically. To begin with, let  $i, j \in \Lambda^n$  and observe that

$$\Delta_{i,j}(t) = r_{i \wedge j}(t) \tilde{\Delta}_{i,j}(t),$$

where, writing  $u, v$  for the sequences obtained from  $i, j$  after deleting the longest initial segment,

$$\tilde{\Delta}_{i,j}(t) = \Delta_{u,v}(t).$$

Differentiating  $p$  times,

$$\begin{aligned} \tilde{\Delta}_{i,j}^{(p)}(t) &= \frac{d^p}{dt^p} (r_{i \wedge j}(t)^{-1} \cdot \Delta_{i,j}(t)) \\ &= \sum_{q=0}^p \binom{p}{q} \cdot \frac{d^q}{dt^q} (r_{i \wedge j}(t)^{-1}) \cdot \Delta_{i,j}^{(p-q)}(t). \end{aligned}$$

A calculation shows that

$$\left| \frac{d^q}{dt^q} (r_{i \wedge j}(t)^{-1}) \right| \leq O_{q, r_{\min}, R_q} (|i \wedge j|^q \cdot r_{i \wedge j}(t)^{-1}).$$

Thus we have the bound

$$|\tilde{\Delta}_{i,j}^{(p)}(t)| = O_{p, r_{\min}, R_p} \left( \max_{0 \leq q \leq p} \left( |i \wedge j|^q \cdot r_{i \wedge j}(t)^{-1} \cdot |\Delta_{i,j}^{(q)}(t)| \right) \right).$$

**Proposition 5.7.** Suppose  $r_i(\cdot), a_i(\cdot)$  are real-analytic on  $I$ . Suppose that for  $i, j \in \Lambda^{\mathbb{N}}$ ,  $\Delta_{i,j} \equiv 0$  on  $I$  if and only if  $i = j$ . Then the associated family  $\{\Phi_t\}_{t \in I}$  is transverse of order  $k$  for some  $k$ .

*Proof.* First, for  $x \in I$  we can extend  $r_i, a_i$  analytically to a complex neighborhood  $U_x$  of  $x$  on which  $|r_i|$  are still bounded uniformly away from 1. Define  $\Delta_{i,j}(z)$  as before for  $i, j \in \Lambda^n$  and  $z \in U_x$ , and note that for  $i, j \in \Lambda^{\mathbb{N}}$  the limit  $\Delta_{i,j}(z) = \lim \Delta_{i_1 \dots i_n, j_1 \dots j_n}(z)$  is uniform for  $z \in U_x$ . This shows that  $\Delta_{i,j}(t)$  is also real-analytic on  $I$ .

Given  $k$ , from the expression for  $\tilde{\Delta}_{i,j}^{(p)}$  above, we see that if  $c > 0$  and there exists  $t_0 \in I$  such that  $|\Delta_{i,j}^{(p)}(t_0)| \leq c \cdot |i \wedge j|^{-p} \cdot r_{i \wedge j}(t_0)$  for all  $0 \leq p \leq k$ , then  $|\tilde{\Delta}_{i,j}^{(p)}(t_0)| \leq c'$  for all  $0 \leq p \leq k$ , where  $c' = O_{k, R_k}(c)$ . For each  $k$  choose  $c_k > 0$  such that the associated  $c'_k$  satisfies  $c'_k < 1/k$ .

Suppose that for all  $k$  the family  $\{\Phi_t\}$  is not transverse of order  $k$ . Then by assumption we can choose  $n(k)$  and distinct  $i^{(k)}, j^{(k)} \in \Lambda^{n(k)}$ , and a point  $t_k \in I$ , such that  $|\Delta_{i^{(k)}, j^{(k)}}^{(p)}(t_k)| \leq c_k \cdot |i^{(k)} \wedge j^{(k)}|^{-p} \cdot r_{i^{(k)} \wedge j^{(k)}}(t_k)$  for  $0 \leq p \leq k$ , and hence  $|\tilde{\Delta}_{i^{(k)}, j^{(k)}}^{(p)}(t_k)| \leq c'_k$ . Let  $u^{(k)}$  and  $v^{(k)}$  denote the sequences obtained from  $i^{(k)}$  and  $j^{(k)}$



by deleting the first  $|i^{(k)} \wedge j^{(k)}|$  symbols, so that the first symbols of  $u^{(k)}$  and  $v^{(k)}$  now differ and  $\Delta_{u^{(k)}, v^{(k)}} = \tilde{\Delta}_{i^{(k)}, j^{(k)}}$ . Hence we have

$$|\Delta_{u^{(k)}, v^{(k)}}^{(p)}(t_k)| \leq c'_k < 1/k \quad \text{for all } 0 \leq p \leq k. \quad (43)$$

Passing to a subsequence  $k_\ell$ , we may assume that  $t_{k_\ell} \rightarrow t_0$  and that  $u^{(k_\ell)} \rightarrow u \in \Lambda^{\mathbb{N}}$  and  $v^{(k_\ell)} \rightarrow v \in \Lambda^{\mathbb{N}}$  (the latter in the sense that all coordinates stabilize eventually to the corresponding coordinate in the limit sequence). Note that  $u \neq v$ , because  $u^{(k_\ell)}, v^{(k_\ell)}$  differ in their first symbol for all  $\ell$ , hence so do  $u, v$ . It follows that  $\Delta_{u^{(k_\ell)}, v^{(k_\ell)}} \rightarrow \Delta_{u, v}$  uniformly and that the same holds for  $p$ -th derivatives. Hence for all  $p \geq 0$ , using uniform convergence and (43),

$$|\Delta_{u, v}^{(p)}(t_0)| = \lim_{\ell \rightarrow \infty} |\Delta_{u^{(k_\ell)}, v^{(k_\ell)}}^{(p)}(t_{k_\ell})| = 0.$$

But  $\Delta_{u, v}$  is real analytic so the vanishing of its derivatives implies  $\Delta_{u, v} \equiv 0$  on  $I$ , contrary to the hypothesis.  $\square$

We turn now to the implications of transversality. The key implication is provided by the following simple lemma.

**Lemma 5.8.** *Let  $k \in \mathbb{N}$  and let  $F$  be a  $k$ -times continuously differentiable function on a compact interval  $J \subseteq \mathbb{R}$ . Let  $M = \|F\|_{J, k}$  and let  $0 < c < 1$  be such that for every  $x \in J$  there is a  $p \in \{0, \dots, k\}$  with  $|F^{(p)}(x)| > c$ . Then for every  $0 < \rho < c/2^k$ , the set  $F^{-1}(-\rho, \rho) \subseteq J$  can be covered by  $O_{k, M, |J|}(1/c^2)$  intervals of length  $\leq 2(\rho/c)^{1/2^k}$  each.*

*Proof.* For brevity, we shall suppress dependence on the parameters  $k, M, |J|$ , so throughout this proof,  $O(\cdot) = O_{k, M, |J|}(\cdot)$ .

The proof is by induction on  $k$ . For  $k = 0$  the hypothesis is that  $|F^{(0)}(x)| = |F(x)| > c$  for all  $x \in J$ , hence  $F^{-1}(-\rho, \rho) = \emptyset$  for  $0 < \rho < c = c/2^0$ , and the assertion is trivial.

Assume that we have proved the claim for  $k - 1$  and consider the case  $k$ . Let  $J'$  be a maximal closed interval in  $F^{-1}[-c, c]$  and let  $G = F'|_{J'}$ . Note that  $G$  satisfies the hypothesis for  $k - 1$  and the same value of  $c$  and  $M$ , and  $\sqrt{c\rho} < c/2^{k-1}$ , so from the induction hypothesis we find that  $G^{-1}(-\sqrt{c\rho}, \sqrt{c\rho})$  can be covered by  $O(1/c)$  intervals of length  $< 2(\sqrt{c\rho}/c)^{1/2^{k-1}} = 2(\rho/c)^{1/2^k}$  each. Let  $U$  denote the union of this cover and consider the intervals  $J'_i$  which are the closures of the maximal sub-intervals in  $J' \setminus U$ . By the above, the number of such intervals  $J'_i$  is  $\leq O(1/c)$ . Now, on each  $J'_i$  we have  $|F'| \geq \sqrt{c\rho}$ , so by continuity of  $F'$  either  $F' \geq \sqrt{c\rho}$  or  $F' \leq -\sqrt{c\rho}$  in all of  $J'_i$ . An elementary consequence of this is that  $J'_i \cap F^{-1}(-\rho, \rho)$  is an interval of length at most  $2\rho/\sqrt{c\rho} = 2\sqrt{\rho/c} \leq 2(\rho/c)^{1/2^k}$ . In summary we have covered  $J' \cap F^{-1}(-\rho, \rho)$  by  $O(1/c)$  intervals of length  $2(\rho/c)^{1/2^k}$  each.

It remains to show that there are  $O(1/c)$  maximal intervals  $J' \subseteq F^{-1}[-c, c]$  as in the paragraph above. In fact, we only need to bound the number of such  $J'$  that intersect  $F^{-1}(-\rho, \rho)$ . For  $J'$  of this kind, if  $J' = J$  we are done, since this means there is just one such interval. Otherwise there is an endpoint  $a \in J'$  with  $|F(a)| = c$ . There is also a point  $b \in J'$  with  $|F(b)| < \rho < c/2^k$ . Since  $|F'| \leq M$ , we conclude that  $|J'| \geq |b - a| \geq (c - \rho)/M \geq c/2M$ . Thus, since the intervals  $J'$  are disjoint, their number is  $\leq |J|/(c/2M) = O(1/c)$ , completing the induction step.  $\square$

Let  $\text{bdim } X$  denote the upper box dimension of a set  $X$ , defined by

$$\text{bdim } X = \limsup_{r \rightarrow 0} \frac{\log \# \min\{\ell : X \text{ can be covered by } \ell \text{ balls of radius } r\}}{\log(1/r)}.$$

One always has  $\dim X \leq \text{bdim } X$ . The packing dimension is defined by

$$\text{pdim } X = \inf_n \left\{ \sup_n \text{bdim } X_n : X \subseteq \bigcup_{n=1}^{\infty} X_n \right\}.$$

Note that  $\dim X \leq \text{pdim } X$ , and  $Y \subseteq X$  implies  $\text{pdim } Y \leq \text{pdim } X$ .

**Theorem 5.9.** *If  $\{\Phi_t\}_{t \in I}$  satisfies transversality of order  $k \geq 1$  on the compact interval  $I$ , then the set  $E$  of “exceptional” parameters in Theorem 1.7 has packing (and hence Hausdorff) dimension 0.*

*Proof.* Write

$$M = \sup_n \sup_{i,j \in \Lambda^n} \|\Delta_{i,j}\|_{I,k}.$$

That  $M < \infty$  follows from  $k$ -fold continuous differentiability of  $r_i(\cdot), a_i(\cdot)$  and the fact that  $|r_i|$  are bounded away from 1 on  $I$ . By transversality there is a constant  $c > 0$  such that for every  $t \in I$ , every  $n$  and all distinct  $i, j \in \Lambda^n$ ,

$$\left| \frac{\partial^p}{\partial t^p} \Delta_{i,j}(t) \right| > c \cdot |i \wedge j|^{-p} \cdot r_{\min}^{|i \wedge j|} \quad \text{for some } p \in \{0, \dots, k\}.$$

In what follows we suppress the dependence on  $k, M, c$  and  $|I|$  in the  $O(\cdot)$  notation:  $O(\cdot) = O_{k,M,c,|I|}(\cdot)$ .

We may assume that  $c < 1$  and  $k \geq 2$ . Let  $\varepsilon < cr_{\min}/2k$  and fix  $n$  and distinct  $i, j \in \Lambda^n$ . By the previous lemma, for all  $0 < \rho < c|i \wedge j|^{-k} r_{\min}^{|i \wedge j|}/2^k$ , and in particular for  $0 < \rho < cr_{\min}^n/(2n)^k$ , the set  $\{t \in I : |\Delta_{i,j}| < \rho\}$  can be covered by at most  $O((2n)^k/r_{\min}^n)$  intervals of length  $2((2n)^k \rho/r_{\min}^n)^{1/2^k}$  each. Now set  $\rho = \varepsilon^n$  (our choice of  $\varepsilon$  guarantees that  $\rho$  is in the proper range) and let  $i, j$  range over their  $\leq |\Lambda|^n$  different possible values. We find that the set

$$E_{\varepsilon,n} = \bigcup_{i,j \in \Lambda^n, i \neq j} (\Delta_{i,j})^{-1}(-\varepsilon^n, \varepsilon^n)$$

can be covered by  $O((2n)^k |\Lambda|^n / r_{\min}^n)$  intervals of length  $\leq ((2n)^k \varepsilon^n / r_{\min}^n)^{1/2^k}$ . Now,  $E \subseteq E_{\varepsilon}$  where

$$E_{\varepsilon} = \bigcup_{N=1}^{\infty} \bigcap_{n > N} E_{\varepsilon,n}. \quad (44)$$

By the above, for each  $\varepsilon$  and  $N$  we have

$$\begin{aligned} \text{bdim} \left( \bigcap_{n > N} E_{\varepsilon,n} \right) &\leq \lim_{n \rightarrow \infty} \frac{\log(O(2n)^k |\Lambda|^n / r_{\min}^n)}{\log(((2n)^k \varepsilon^n / r_{\min}^n)^{1/2^k})} \\ &= O(2^k \frac{\log(|\Lambda|/r_{\min})}{\log(\varepsilon/r_{\min})}). \end{aligned}$$

The last expression is  $o(1)$  as  $\varepsilon \rightarrow 0$ , uniformly in  $N$ . Thus by (44), the same is true of  $E_{\varepsilon}$ , and  $E \subseteq E_{\varepsilon}$  for all  $\varepsilon$ , so  $E$  has packing (and Hausdorff) dimension 0.  $\square$

Theorem 1.8 now follows by combining Proposition 5.7 and Theorem 5.9.

## 5.5 Miscellaneous proofs

To complete the proof of Corollary 1.5 we have:

**Lemma 5.10.** *Let  $A \subseteq \mathbb{R}$  be a finite set of algebraic numbers over  $\mathbb{Q}$ . Then there is a constant  $0 < s < 1$  such that any polynomial expression  $x$  of degree  $n$  in the elements of  $A$ , either  $x = 0$  or  $|x| > s^n$ .*

*Proof.* Choose an algebraic integer  $\alpha$  such that  $A \subseteq \mathbb{Q}(\alpha)$ . Since the statement is unchanged if we multiply all elements of  $A$  by an integer, we can assume that the elements of  $A$  are integer polynomials in  $\alpha$  of degree  $\leq d$  and coefficients bounded by  $N$ , for some  $q, N$ . Substituting these polynomials into the expression for  $x$ , we have an expression  $x = \sum_{k=0}^{qn} n_k \alpha^k$  where  $n_k \in \mathbb{N}$  and  $|n_k| \leq N$ . It suffices to prove that any such expression is either 0 or  $\geq s^n$  for  $0 < s < 1$  independent of  $n$  (but depending on  $\alpha, N$ ). In proving this last statement we may assume that  $q = 1$  (replace  $s$  by  $s^{1/q}$ ).

Let  $\alpha = \alpha_1, \alpha_2, \dots, \alpha_d$  denote the algebraic conjugates of  $\alpha$  and  $\sigma_1, \sigma_2, \dots, \sigma_d$  the automorphisms of  $\mathbb{Q}(\alpha)$ , with  $\sigma_i \alpha = \alpha_i$ . If  $x \neq 0$  then  $\prod_{i=1}^d \sigma_i(x) \in \mathbb{Z}$ , so

$$1 \leq \left| \prod_{i=1}^d \sigma_i(x) \right| = x \cdot \prod_{i=2}^d \left| \sum_{k=0}^n n_k \sigma_i(x)^k \right| \leq x \cdot \prod_{i=2}^d \sum_{k=0}^n n_k |\alpha_i|^k \leq x \cdot (n \cdot N \cdot \alpha_{\max}^n)^d,$$

where  $\alpha_{\max} = \max\{|\alpha_2|, \dots, |\alpha_d|\}$ . Dividing out gives the lemma.  $\square$

We finish with some comments on Sinai's problem, Theorem 1.11. We first state a generalization of Theorem 1.7 needed to treat families of IFSs that contract only on average.

Suppose that for  $t \in I$  we have a family  $\Phi_t = \{\varphi_{i,t}\}_{i \in \Lambda}$  of (not necessarily contracting) similarities of  $\mathbb{R}$ , and as usual write  $\varphi_{i,t} = r_{i,t}U_{i,t} + a_{i,t}$ . Let  $p$  be a fixed probability vector and suppose that for each  $t$  we have  $\sum p'_i \log r_i < 0$ , i.e. the systems contract on average. One can then show that there is a unique probability measure  $\mu_t$  on  $\mathbb{R}$  satisfying  $\mu_t = \sum_{i \in \Lambda} p_i \cdot \varphi_{i,t} \mu_t$  [20], that  $H(\mu_t, \mathcal{D}_m) < \infty$  for every  $t$  and  $m$ , and that  $\mu_t([-R, R]) \rightarrow 1$  as  $R \rightarrow \infty$  uniformly in  $t$ . Under these conditions one can verify the stronger property that for every  $t \in I$  we have

$$|H_m(\mu_t) - H_m((\mu_t)_{[-R, R]})| = o(1) \quad \text{as } R \rightarrow \infty$$

uniformly in  $t$  and  $m$ .

**Theorem 5.11.** *Let  $(\Phi_t)_{t \in I}$ ,  $p$ , and  $\mu_t$  be as in the preceding paragraph. Let  $\tilde{\mu}$  denote the product measure on  $\Lambda^{\mathbb{N}}$  with marginal  $p$ , and suppose that  $A \subseteq \Lambda^{\mathbb{N}}$  is a Borel set such that  $\tilde{\mu}(A) > 0$ . Write*

$$E = \bigcap_{\varepsilon > 0} \left( \bigcup_{N=1}^{\infty} \bigcap_{n > N} \left( \bigcup_{i,j \in A} (\Delta_{i,j})^{-1}((-\varepsilon^n, \varepsilon^n)) \right) \right).$$

*Then  $\dim \mu_t = \min\{d, \text{s-dim } \mu_t\}$  for every  $t \in I \setminus E$ . Furthermore suppose that  $I \subseteq \mathbb{R}$  is compact and connected, and that the parametrization is analytic in the sense of Theorem 1.8. If*

$$\forall i, j \in A \quad (\Delta_{i,j} \equiv 0 \text{ on } I \iff i = j)$$

*then the set  $E$  above is of packing (and Hausdorff) dimension at most  $k - 1$ , and in particular of Lebesgue measure 0.*

The proof is the same as the proofs of Theorems 1.7 and 1.8, except that in analyzing the resulting convolution one must approximate  $\mu_t$  by  $(\mu_t)_{[-R,R]}$  for an appropriately large  $R$  that is fixed in advance, with the scale  $n$  large relative to  $R$ . We omit the details.

Let us see how this applies to Theorem 1.11, where  $\varphi_{-1,\alpha}(x) = (1 - \alpha)x - 1$  and  $\varphi_{1,\alpha}(x) = (1 + \alpha)x + 1$  for  $\alpha \in (0, 1]$ , and  $p = (1/2, 1/2)$ . It suffices to consider the system for  $\alpha \in [s, 1]$  for some  $s > 0$ . Let  $A$  be the set of  $i \in \Lambda^{\mathbb{N}}$  such that  $|\frac{1}{N} \sum_{n=1}^N i_n - \frac{1}{2}| < \delta$  for  $n > N(\delta)$ , where  $\delta > 0$  small enough to ensure that  $|\varphi_{i_1 \dots i_n}| < 1$  when this condition holds, and  $N(\delta)$  large enough that  $\tilde{\mu}(A) > 0$ ; in fact we can make  $\tilde{\mu}(A)$  arbitrarily close to 1, by the law of large numbers. It remains to verify for  $i, j \in A$  that  $\Delta_{i,j}$  vanishes on  $[s, 1]$  if and only if  $i = j$ . Note that for  $i \in \{-1, 1\}^n$ ,

$$\varphi_{i,\alpha}(0) = 1 + (1 + i_1\alpha) + (1 + i_1\alpha)(1 + i_2\alpha) + \dots + \prod_{k=1}^n (1 + i_k\alpha).$$

Thus  $\Delta_{i,j}$  is a series whose terms are of the form  $c_{k,m}(1 - \alpha)^k(1 + \alpha)^m$  for some  $c_{k,m} \in \{0, \pm 1\}$ , and  $i = j$  if and only if all terms are 0. Furthermore, there is an  $n_0$  such that if  $k + m \geq n_0$  and  $c_{k,m} \neq 0$ , then  $k > (1 - \delta)m$ . Thus since  $s \leq \alpha \leq 1$  and  $\delta$  was chosen small enough, the series converges uniformly on  $[s, 1]$ , and furthermore there is an  $\varepsilon > 0$  such that the series converges uniformly on some larger interval  $[s, 1 + \varepsilon]$ , and even in a neighborhood of 1 in the complex plane. Hence  $\Delta_{i,j}(\cdot)$  is real-analytic on  $[s, 1 + \varepsilon]$  and is given by this series. Now, if  $i \neq j$  we can divide out by the highest power  $(1 - \alpha)^{k_0}$  that is common to all the terms (possibly  $k_0 = 0$ ), and evaluate the resulting function at  $\alpha = 1$ . We get a finite sum of the form  $\sum_{(k,m) \in U} c_{m,k} 2^m$  for some finite set of indices  $U \subset \mathbb{N}^2$  such that  $c_{m,k} \in \{\pm 1\}$  for  $(k, m) \in U$ . Such a sum cannot vanish, hence by analyticity  $\Delta_{i,j} \neq 0$  on every sub-interval of  $[s, 1 + \varepsilon]$  as desired, and in particular  $\Delta_{i,j} \neq 0$  on  $[s, 1]$ , as desired.

## References

- [1] J. C. Alexander and J. A. Yorke. Fat baker's transformations. *Ergodic Theory Dynam. Systems*, 4(1):1–23, 1984.
- [2] J. Bourgain. On the Erdős-Volkmann and Katz-Tao ring conjectures. *Geom. Funct. Anal.*, 13(2):334–365, 2003.
- [3] J. Bourgain. The discretized sum-product and projection theorems. *J. Anal. Math.*, 112:193–236, 2010.
- [4] J. Bourgain, N. Katz, and T. Tao. A sum-product estimate in finite fields, and applications. *Geom. Funct. Anal.*, 14(1):27–57, 2004.
- [5] Paul Erdős. On a family of symmetric Bernoulli convolutions. *Amer. J. Math.*, 61:974–976, 1939.
- [6] Paul Erdős. On the smoothness properties of a family of Bernoulli convolutions. *Amer. J. Math.*, 62:180–186, 1940.
- [7] Carl-Gustav Esseen. On the Liapounoff limit of error in the theory of probability. *Ark. Mat. Astr. Fys.*, 28A(9):19, 1942.

- [8] De-Jun Feng and Huyi Hu. Dimension theory of iterated function systems. *Comm. Pure Appl. Math.*, 62(11):1435–1500, 2009.
- [9] Harry Furstenberg. Intersections of Cantor sets and transversality of semigroups. In *Problems in analysis (Sympos. Salomon Bochner, Princeton Univ., Princeton, N.J., 1969)*, pages 41–59. Princeton Univ. Press, Princeton, N.J., 1970.
- [10] Hillel Furstenberg. Ergodic fractal measures and dimension conservation. *Ergodic Theory Dynam. Systems*, 28(2):405–422, 2008.
- [11] Adriano M. Garsia. Arithmetic properties of Bernoulli convolutions. *Trans. Amer. Math. Soc.*, 102:409–432, 1962.
- [12] Adriano M. Garsia. Entropy and singularity of infinite convolutions. *Pacific J. Math.*, 13:1159–1169, 1963.
- [13] Michael Hochman. Self-similar sets with overlaps and sumset phenomena for entropy, the multidimensional case. *preprint*, 2012.
- [14] Michael Hochman and Pablo Shmerkin. Equidistribution from fractals. *preprint*, 2011.
- [15] V. A. Kaĭmanovich and A. M. Vershik. Random walks on discrete groups: boundary and entropy. *Ann. Probab.*, 11(3):457–490, 1983.
- [16] Nets Hawk Katz and Terence Tao. Some connections between Falconer’s distance set conjecture and sets of Furstenburg type. *New York J. Math.*, 7:149–187 (electronic), 2001.
- [17] Mike Keane, Meir Smorodinsky, and Boris Solomyak. On the morphology of  $\gamma$ -expansions with deleted digits. *Trans. Amer. Math. Soc.*, 347(3):955–966, 1995.
- [18] Richard Kenyon. Projecting the one-dimensional Sierpinski gasket. *Israel J. Math.*, 97:221–238, 1997.
- [19] T. W. Körner. Hausdorff dimension of sums of sets with themselves. *Studia Math.*, 188(3):287–295, 2008.
- [20] Matthew Nicol, Nikita Sidorov, and David Broomhead. On the fine structure of stationary measures in systems which contract-on-average. *J. Theoret. Probab.*, 15(3):715–730, 2002.
- [21] Yuval Peres and Wilhelm Schlag. Smoothness of projections, Bernoulli convolutions, and the dimension of exceptions. *Duke Math. J.*, 102(2):193–251, 2000.
- [22] Yuval Peres, Wilhelm Schlag, and Boris Solomyak. Sixty years of Bernoulli convolutions. In *Fractal geometry and stochastics, II (Greifswald/Koserow, 1998)*, volume 46 of *Progr. Probab.*, pages 39–65. Birkhäuser, Basel, 2000.
- [23] Yuval Peres, Károly Simon, and Boris Solomyak. Absolute continuity for random iterated function systems with overlaps. *J. London Math. Soc. (2)*, 74(3):739–756, 2006.

- [24] Yuval Peres and Boris Solomyak. Problems on self-similar sets and self-affine sets: an update. In *Fractal geometry and stochastics, II (Greifswald/Koserow, 1998)*, volume 46 of *Progr. Probab.*, pages 95–106. Birkhäuser, Basel, 2000.
- [25] Mark Pollicott and Károly Simon. The Hausdorff dimension of  $\lambda$ -expansions with deleted digits. *Trans. Amer. Math. Soc.*, 347(3):967–983, 1995.
- [26] F. Przytycki and M. Urbański. On the Hausdorff dimension of some fractal sets. *Studia Math.*, 93(2):155–186, 1989.
- [27] Jörg Schmeling and Pablo Shmerkin. On the dimension of iterated sumsets. In *Recent developments in fractals and related fields*, Appl. Numer. Harmon. Anal., pages 55–72. Birkhäuser Boston Inc., Boston, MA, 2010.
- [28] Pablo Shmerkin and Boris Solomyak. Zeros of  $\{-1, 0, 1\}$  power series and connectedness loci for self-affine sets. *Experiment. Math.*, 15(4):499–511, 2006.
- [29] Boris Solomyak. On the random series  $\sum \pm \lambda^n$  (an Erdős problem). *Ann. of Math. (2)*, 142(3):611–625, 1995.
- [30] Grzegorz Świątek and J. J. P. Veerman. On a conjecture of Furstenberg. *Israel J. Math.*, 130:145–155, 2002.
- [31] Terence Tao. Sumset and inverse sumset theory for Shannon entropy. *Combin. Probab. Comput.*, 19(4):603–639, 2010.
- [32] Terence Tao and Van Vu. *Additive combinatorics*, volume 105 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2006.

Email: mhochman@math.huji.ac.il

Address: Einstein Institute of Mathematics, Givat Ram, Jerusalem 91904, Israel